Author's Note

I wrote this essay in the spring and summer of 2014, when I submitted it as my thesis for the BPhil in Philosophy at Oxford. I intended to revise parts of it for publication, but life gets in the way. I am therefore making public the original version—typos and all!—hoping it may be useful to philosophers and others interested in homotopy type theory, category theory, dependent type theory, and mathematical structuralism.

If I were to revisit this material, I would no doubt do things differently. Let me briefly sketch how my views had changed, when last I thought about these issues.

The original idea of the thesis was to make the best sense I could of the idea that homotopy type theory provides an 'object structuralist' foundation for mathematics: a theory on which there is a domain of mathematical objects that only have structural (i.e. isomorphism-invariant) properties. The key to this interpretation is obviously the *univalence* axiom, which appears to assert that isomorphic (or, more generally, equivalent) types are identical. Combine this with the thought that identical objects have all the same properties, in the form of the *path-induction* axiom. It follows that all the properties of types are structural. At least, that is the rough idea; the apparent miracle of homotopy type theory is that it gives a framework in which this idea can flourish.

In contrast to 'object structuralism', the metaphysically more modest view of 'property structuralism' says that, whatever sorts of objects there may be, mathematics is (generally speaking) only concerned with their structural properties. I later realised that there is an alternative interpretation of the univalence axiom along property-structuralist lines, as long as we are willing to be slightly less deferential to notation. I defended this interpretation in a few talks in 2015. To indicate the rough idea, we could linguistically do away with the identity relation between types, and instead understand univalence as attributing the path-induction property to the isomorphism relation. (My criticism of related ideas in section 7.2 of chapter IV now strikes me as mistaken.) In classical terms, the upshot would still be that 'all the properties of types are structural', but the trick is that, having set aside the identity relation, we can understand the 'all' as a restricted quantifier, ranging over structural properties only. This is the sort of quantifier restriction that a property structuralist might recommend.

While I am still drawn to the romance of object-structuralism, it seems to me that the property-structuralist interpretation is the more prudent bet, while remaining broadly sympathetic to homotopy type theory, including univalence, as a foundational project. Indeed, it still strikes me as a worthy goal to provide a property-structuralist foundation for mathematics, and it is still remarkable that homotopy type theory is in with a fighting chance. Were I to revise the essay, I would wish to develop and emphasise these points more clearly.

Teruji Thomas
Oxford, 3 April 2021

Thomas, Teruji (2014). *Homotopy Type Theory and Structuralism*. BPhil Thesis, University of Oxford.

Homotopy Type Theory and Structuralism

Teruji Thomas

This essay satisfied the thesis requirement of the
BPhil in Philosophy, 2012–2014

Candidate Number: 352680
Word Count: 29667

ABSTRACT. I explore the possibility of a structuralist interpretation of homotopy type theory (HOTT) as a foundation for mathematics. There are two main aspects to HOTT's structuralist credentials. First, it builds on categorical set theory (CST), of which the best-known variant is Lawvere's ETCS. I argue that CST has merit as a structuralist foundation, in that it ascribes only structural properties to typical mathematical objects. However, I also argue that this success depends on the adoption of a strict typing system which undermines the metaphysical seriousness of this structuralism. Homotopy type theory adds to CST a distinctive theory of identity between sets, which arguably allows its objects to be seen as ante rem structures. I examine the prospects for such a view, and address many other interpretive problems as they arise.

ACKNOWLEDGEMENTS. I thank David Wallace, who supervised this thesis with skill and patience. I thank Neil Dewar and Katherine Robertson for valuable feedback and priceless encouragement. I am also grateful to Kobi Kremnitzer and the other participants in the homotopy type theory reading group that ran in Oxford this last year. I first encountered some of the philosophical literature considered here in conversation with Hilary Greaves and Maria Lasonen-Aarnio; that is the least thing I have to thank them for. Finally, Samantha Johnson and Ryu Thomas gave up a great deal to make this thesis and my other work for the BPhil possible. I dedicate it to them.

Contents

CHAPTER I

Introduction

This thesis represents perhaps the first attempt at a sustained philosophical interpretation of homotopy type theory (HoTT). Over the last decade, HoTT has developed as a major new candidate for mathematical foundations. Unusually for a foundational programme, it has some influential advocates within the mainstream mathematical community. The precondition for this thesis is the recent publication of a textbook (Univalent Foundations Program (2013), henceforth *UFP*), which provides a canonical reference-point for a quickly-evolving field. The guiding light of this thesis is the unavoidable impression that HoTT articulates a distinctively *structuralist* vision of mathematics. For example, Awodey (2014) writes that, according to HoTT,

> mathematical objects simply are structures. Could there be a stronger formulation of structuralism? (p. 11)

My basic goal is to explore and evaluate this claim.[1] But HoTT has so many novel features that I will have to grapple with its interpretation more broadly.

A few words of motivation. If Awodey's conclusion is correct, homotopy type theory may be the closest thing we have to a rigorous articulation of ontological structuralism. The competition is not stiff. Even seeing

[1]Awodey himself does not really argue for it. The closest thing to an argument is that, according to HoTT, '[T]wo mathematical objects are identical if and only if they have the same structure.... In other words, mathematical objects simply are structures.' Accepting the first sentence, the second doesn't follow. To give a mathematical example, two sets X and Y are identical if and only if they have the same singletons $\{X\}$ and $\{Y\}$; it doesn't follow that sets simply are singletons.

how this articulation fails is likely to advance our understanding. Second, nowhere near enough philosophical attention has been paid to the areas of mathematics to which homotopy type theory is particularly adapted: homotopy theory, higher category theory, and their kin. These are not niche areas in twenty-first century mathematics, or even in twenty-first century theoretical physics. But they provide distinctive challenges to the philosophy of mathematics in general and to structuralism in particular. Homotopy type theory claims to provide a newly perspicacious treatment of these fields. That alone would make it worth our while.

In this introduction, I disentangle some of the features of HoTT that are relevant to its foundational aspirations. This will help to clarify both what HoTT is about, and the kind of questions that will interest me here. Then I will briefly lay out some ideas about structuralism that will frame my analysis. Then I will outline the contents of this thesis.

1. Homotopy Type Theory

The first observation is that HoTT has claims both as a foundation for *mathematics in general* and as a foundation for *homotopy theory in particular*. Let me begin with the latter. On a first pass, one can think of the objects of homotopy theory as sets with a certain kind of structure, much as groups or manifolds are sets with structure. (On a second pass, these objects are simply topological spaces, but the 'structure-preserving maps' between them are not simply the continuous functions.) On one reading, the objects of HoTT — the eponymous *types* — just are those sets with structure, and HoTT is a theory about them.[2] So, on that reading, HoTT is only a 'foundation' for a proper part of standard mathematics, in roughly the

[2]It is important to realise that in this heuristic description I am using the word 'theory' in a different sense from the usual model-theoretic one. HoTT comprises not only axioms but also a whole deductive system.

way that Euclid's postulates sought to provide a foundation for geometry. The basic reason to seek such a foundation is a sense — developed over decades of mathematical practice — that the classical set-theoretic treatment of homotopy theory, while perhaps formally adequate, fails to make perspicuous the main concepts and results.

Adopting a new framework like HoTT leaves open a range of possible interpretations. A conservative interpretation is that HoTT is merely a 'fresh look' at homotopy theory, which is ultimately to be interpreted in familiar set-theoretic terms. In analogy, the adoption of 'line' as a primitive in the development of Euclidean geometry does not preclude the view that straight lines are sets of points. It is simply more useful, for certain purposes, to start from the high-level conceptual framework employed by the postulates.

A radical alternative is to think of HoTT as a 'fresh start'. On this view, HoTT is no more to be interpreted in some background set theory, like ZFC, than set theory itself is to be interpreted in such a background. The objects of HoTT can be *represented* and studied within ZFC, as the standard development of homotopy theory indicates. But this only attests to the representational power of set theory, not to the conceptual or ontological priority of sets over homotopy types.

This radical alternative is encouraged by the observation that most of ordinary mathematics (including traditional homotopy theory) can be developed *inside* HoTT, using what amount to the standard set-theoretic techniques. This is because sets are *among* the objects of homotopy theory: in terms of the standard development of the subject, every set has a default ('discrete') homotopy structure. In this sense, HoTT has potential as a foundation for mathematics in general. On the face of it, HoTT even generalises standard mathematics, since only some types are sets.

As a set-theoretic foundation, however, HoTT has two distinctive features. The first feature is that there is no membership relation between sets. For this reason, the sets are not governed by Zermelo set theory (ZST), but rather by a version of categorical set theory (CST).[3] This latter set theory has played a prominent role in philosophical discussions of 'category-theoretic structuralism.' In particular, it has been alleged that CST is a distinctively structuralist foundation, in contrast to ZST.

The second feature is the subtle treatment of *identity*. On the face of it, HoTT claims that all sets of the same cardinality are identical, and in somewhat more generality that isomorphic mathematical systems are identical. This is the fundamental *Univalence* axiom; it is what led Awodey to declare that 'mathematical objects simply are structures.'

This sketch suggests two distinct (although not wholly independent) lines of inquiry for a philosophical interpretation and evaluation of HoTT. First, what should one make of the apparent generalisation of set theory to homotopy theory? Second, what should one make of the treatment of conventional mathematics within the set-theoretic fragment of HoTT? In this thesis I am going to focus primarily on the second question. In particular, I will examine the prospects for a structuralist interpretation of homotopy type theory.

To complete this introduction to the foundational claims of homotopy type theory, I should mention two closely linked features of HoTT that are

[3]I use 'Zermelo set theory' to refer generically to Z, ZFC, and whatever else, although I will usually assume there are no ur-elements. Similarly, the canonical version of categorical set theory is Lawvere's ETCS(Lawvere, 1964). However, there are many possible extensions and reformulations which I include under this banner.

sometimes put forward as advantages of HoTT over conventional foundations, but which I intend largely to set aside. These features are *automation* and *constructivity*.

On the first point, HoTT lends itself to automated proof-checking and (at least partly) automated proof-generation. This is not a trivial concern, if one thinks that such automation will play a significant role in future mathematical practice. It is also potentially a 'foundational' concern, to the extent that 'foundations' ought to codify, ease, and shed light upon such practice. Be that as it may, I am interested in more metaphysical questions. For example, how does the proposed foundation relate to the *subject matter* of mathematics, if there is such a thing? I do not think that computation and automation are immediately relevant to such questions.

As for the second feature I mentioned, the formal system of HoTT is a version of Martin-Löf's intuitionistic type theory (Martin-Löf, 1984). However, in so far as it matters, I will enforce classical logic. What this means is a bit complicated, because there turn out to be several possible views about what 'logic' is in HoTT. For the cognoscenti: I will henceforth assume the axiom of choice for sets, which entails LEM for mere propositions as well as other useful results like propositional resizing (*UFP*, ch. 3). At a technical level, the adoption of classical logic will not interfere with the other features of HoTT that interest me. At a philosophical level, two basic points can be made. First, intuitionistic logic is typically understood as a logic of provability or constructibility, rather than as a logic of *truth*. For that reason, LEM is a natural part of broadly realist philosophies of mathematics. Realism is a natural default for this kind of interpretive project: it allows me to ask what HoTT claims about the world. The second point is that, regardless of whether one is a realist, it is worth asking what sort of framework HoTT provides for *mainstream* — hence non-constructive — mathematics.

Two serious caveats must be added to this endorsement of classical logic. They will be important later on. First, HoTT is, as the name suggests, a kind of type theory, and the class of well-typed sentences is more limited than one might expect. For example, there is no way to express equality between elements of distinct sets. There is then no question of whether such equalities satisfy the law of excluded middle. Second, as already advertised, HoTT has a distinctive treatment of equality between sets. On one reading, this means that the inference rules for identity are broadly intuitionistic. I will reject this reading, but it is a major issue we must face.

2. Property Versus Object Structuralism

The question that guides my approach to HoTT is whether and how it articulates a structuralist vision of mathematics. I will assume that the reader has a passing familiarity with the basic motivations and platitudes of mathematical structuralism. I will therefore introduce ideas that seem relevant as I go, and avoid a systematic treatise. But let me draw here the basic distinction that will frame my analysis. This is the distinction between *property* and *object* structuralism.[4]

For concreteness, start with standard Zermelo set theory. To get the ball rolling, I am happy to grant that Zermelo set theory is true and epistemically secure. Now, we can find many models of, say, the second-order Peano axioms among the Zermelo sets. These models might *represent* the

[4]The terminology isn't standard. It lines up roughly with the more common eliminivist/non-eliminitivist distinction (e.g Parsons, 1990; Cole, 2010), but I frame things in a way particularly relevant to my current needs. Perhaps the most useful overview and classificatory discussion is in (Reck and Price, 2000).

natural numbers, but no one of them *is* the natural numbers. That is what Benacerraf's famous argument (Benacerraf, 1965) seems to show. If Zermelo set theory is a foundation for mathematics, it isn't because things like the natural numbers are among its objects. But that isn't the only way in which Zermelo set theory might be foundational. The basic structuralist point of view is that number theory is about the structure that all these models have in common. Zermelo set theory could still be 'foundational' in that it allows us to get at this structure. (Note I will thus always distinguish between 'models' or 'systems' on the one hand and 'structures' on the other; the former 'exemplify' or 'instantiate' the latter.)

What sort of thing is this structure? We might start by talking about the structural properties of the various models. Theorems of of arithmetic express such properties. A model of the Peano axioms consists of a set X, a function $f : X \to X$, and an element $0 \in X$. One axiom of arithmetic is the sentence

$$\neg\exists(x \in X).(0 = f(x)). \tag{1}$$

This expresses a relation between X, f, and 0, and so a property of the system. This property is instantiated whenever X, f, and 0 form a model of the natural numbers. It is a structural property.

At a first pass, then, the structure of the natural numbers is just the conjunction of the structural properties. By *property structuralism* I mean the view that mathematics is about set-theoretic models and their structural properties. According to property structuralism, Zermelo set theory is not an ideal foundation for mathematics. It is not a ideal foundation because the properties of systems that are definable in Zermelo set theory are not always structural. For example, 'containing $\{\{\emptyset\}\}$' is not a structural property of natural numbers systems. Some models contain $\{\{\emptyset\}\}$, and others do not. Of course, that doesn't mean that one can't get along using Zermelo set theory. One just has to be careful not to dwell on non-structural properties.

One way to be careful, formally, would be to devise a language in which only structural properties can be expressed. Theories expressed in such a language would attribute to their objects only structural properties. The objects might, to be sure, have other properties as well. But, according to property structuralism, they simply aren't part of mathematics. I mentioned that an important aspect of homotopy type theory is its relationship with categorical set theory. One of my main arguments will be that a certain version of categorical set theory provides this kind of property-structuralist language.

Property structuralism doesn't give a simple account of what we are doing when we talk about 'the natural numbers'. When I say 'zero is not a successor', I could be doing one of (at least) two things. I could be saying of some particular model $(X, f, \mathtt{0})$ that (1) is true. But that particular model has no particular claim to be called 'the natural numbers'. Or I could be making a universally quantified statement that (1) is true of *all* models of the natural numbers. But, still, there is no particular thing that is the natural numbers. The very natural numbers are eliminated, either in favour of some particular (but basically arbitrary) model, or by means of universal quantification. We are still not talking about *the natural numbers*. This is an important objection to property structuralism. I will return to it in Chapter III.

In contrast to property structuralism, what I will call *object structuralism* seeks to preserve the natural numbers as a genuine object of study. This object is an 'ante rem structure' in the terminology of Shapiro (1997).[5] In particular, the ante rem structure of the natural numbers is the intended model of arithmetic. This model is not among the Zermelo sets. But Zermelo set theory could still be a foundation for object structuralism. We could still get at the natural numbers themselves by abstracting from the models of arithmetic in Zermelo set theory. It would again be incumbent upon us to prescind from the non-structural features of Zermelo sets. Obviously, though, Zermelo set theory cannot be an ideal object-structuralist foundation, because it does not directly describe ante rem structures.[6]

If object structuralism is right, it would be more perspicacious to replace the formal theory of Zermelo sets by a formal theory of ante rem structures. On the face of it, Shapiro suggested a way to do this in his book (Shapiro, 1997, ch. 3). But, as Shapiro himself emphasised, this was just a superficial reworking of set theory. If Awodey is right that in homotopy type theory 'mathematical objects simply are structures', then perhaps HoTT is a better formalisation of object structuralism. That is what we have to see.

It must be emphasised that there are serious objections to the cogency of object structuralism. Many of them are summarised in (MacBride, 2005). I will touch on some of these objections as we go, but it is not my business

[5]I will continue to use the phrase 'ante rem structure' because, like Shapiro's, each of the structures is supposed to come along with a system of places that themselves exemplify the structure. But I will not pay much attention to the question of ontological priority suggested by the phrase 'ante rem.'

[6]Shapiro (Shapiro, 2005) explains a slightly different way in which Zermelo set theory could play a foundational role in structuralist mathematics. We might try to specify a structure axiomatically, but our success in this will depend on whether the axioms are consistent or 'coherent' in some more general sense. Model theory gives us a way to study this coherence.

here to seriously debate the merits of object structuralism. On the other hand, it would represent some progress to see how homotopy type theory articulates, or fails to articulate, a version of the object-structuralist view.

3. Prospectus

Now let me turn to an outline of this paper. I will develop and examine the ideas behind homotopy type theory in three stages.

In the first stage (Chapter II), I consider categorical set theory. As I have mentioned, HoTT builds upon CST, which has its own claims as a structuralist foundation. My aim here is two-fold. First, I want to lay groundwork for the discussion of HoTT. Second, I want to evaluate CST on its own terms. The baseline sense in which CST is 'structuralist' has, I think, been correctly diagnosed by McLarty (2004). It embodies property structuralism. I consider, and reject, some epistemic criticisms of CST as a foundation. But my main argument is that property structuralism places strong constraints on the formal development of the theory. In particular, it seems that we must abandon identity predicates between sets, and this in turn requires the use of a language with 'dependent types'.

Such a language is the heart of Martin-Löf's intuitionistic type theory (MLTT), which I discuss in the second stage (Chapter III). I make clear how MLTT — with set identities omitted[7] — is naturally viewed as a development of CST along the lines drawn in Chapter II. However, it also introduces some new features. The new feature most *immediately* relevant to the structuralist interpretation is the use of names for specific structurally-characterised types. For example, MLTT appears to name a specific set of

[7]There are two standard flavours of MLTT — 'extensional' and 'intensional' — distinguished by their treatment of identities between sets. Homotopy type theory ultimately develops the intensional version, but, at this stage in the argument, I want simply to omit set identities.

natural numbers, whereas CST, as I present it, posits the mere existence of an appropriate system. At first sight, this suggests that MLTT has made the leap from property structuralism to *object structuralism*. However, I argue that the object-structuralist interpretation may not be mandatory at this stage, as long as we attend carefully to the semantic role of singular terms in a property-structuralist language.

In the third stage (Chapter IV) I explain how homotopy type theory builds upon my identity-free version of Martin-Löf type theory by reintroducing identities between sets. Crucially, this identity predicate satisfies a *skeletal axiom*, standardly known as *Univalence*. It says that equinumerous sets are equal, and entails in some generality that isomorphic systems are equal. It is this feature that led Awodey to describe HoTT as an articulation of structuralism.

There is a surface reading of HoTT according to which the inference rules for equality are broadly intuitionistic. I will explain how to reject this view, although this rejection leaves open a number of interpretive issues. The biggest of these is that the skeletal axiom forces typing to be *intensional*: even though sets S and T may be equal, there may be no well-typed equations between elements of S and elements of T. How should we interpret this? I suggest an interpretation that arises fairly naturally from an object-structuralist point of view. This strikes me as the best hope for a structuralist interpretation of homotopy type theory. But I also consider the view that the set-identity predicate of homotopy type theory does not express identity at all.

In Chapter V, I give a preliminary discussion of some problems for future work. First, I consider the question of when models of different theories have the same structure. Does homotopy type theory help us here?

Second, I consider the way in which HoTT generalises set-theory to homotopy theory, and speculate about the implications of this generalisation for the structuralist interpretation.

Chapter VI is the conclusion.

CHAPTER II

Categorical Set Theory

As I explained in the introduction, some of the structuralist flavour of HoTT comes from its close relation to categorical set theory. The best-known form of CST is Lawvere's ETCS, the elementary theory of the category of sets. I will discuss a version of it here.

The basic difference between Zermelo set theory and categorical set theory is that the former takes as primitive the idea of one set being a member of another, while the latter takes as primitive the idea of a function from one set to another. In section 5 of this chapter, I will discuss the relative priority of these approaches, but, for now, let us examine their effects.

In Zermelo set theory, sets are characterised by their extensions. For example, 'the set of natural numbers' might be defined inductively by

$$0 :\equiv \emptyset \text{ and } (n+1) :\equiv \{n\} \qquad \text{or} \qquad 0 :\equiv \emptyset \text{ and } (n+1) :\equiv n \cup \{n\}$$

or by some similar device. Different such definitions attribute to the natural numbers different properties: on the first definition, for example, 3 is an element of 4, but not on the second. These properties that vary from one model to another are non-structural. Moreover, these non-structural properties seem inappropriate to the natural numbers. It is a category mistake to ask after the elements of the number 4.

In contrast, categorical set theory takes as primitive the notion of *function*. Each function has a domain and codomain; functions can be composed. This language of functions, domains, codomains, and composition, is the language of *category theory*.[1] Categorical set theory is the theory, in that language, of a particular category, the category of sets. Roughly speaking, this means that in CST, we can only characterise the set of natural numbers in terms of the functional relations in which it stands. Such relations are inevitably structural. So, at least, say the basic platitudes of *category-theoretic structuralism.*

My goal in this chapter is to look more closely at these platitudes. The main argument is that CST has potential as a property-structuralist foundation (see ch. I.2), but this potential can only be fulfilled if category theory is formulated in a language with strict type distinctions. That conclusion will motivate, in the next chapter, the formalism of dependent type theory. In §5 I return to the issue of whether CST depends metaphysically and epistemically on Zermelo set theory.

1. The Two-Sorted Theory of Categories

Let me start with a quick exposition of the standard theory of categories. A category involves 'objects' and 'morphisms' between them. The preeminent example is the category of *sets* (as objects) and *functions* (as morphisms). As one might expect from this example, each morphism f has a domain $d(F)$ and a codomain $c(F)$, which are objects. If $d(f) = c(g)$ then f and g can be composed to form a new morphism $f \circ g$. And each object X has an identity morphism id_X.

[1]One theme of this chapter is that there are several languages in which one can formulate category theory — I will be interested in which one is best for structuralism.

Following this sketch, the theory of categories is usually presented as a theory in a first-order language with two sorts, $\mathscr{O}$ for objects and $\mathscr{M}$ for morphisms,[2] along with four function symbols

$$d \qquad c \qquad \mathrm{id} \qquad \circ$$

The language also contains identity predicates on $\mathscr{O}$ and on $\mathscr{M}$. More precisely 'd' and 'c' are function symbols of type $\mathscr{M} \to \mathscr{O}$, and 'id' has type $\mathscr{O} \to \mathscr{M}$. On the other hand, '$\circ$' is to represent a *partially defined* function $\mathscr{M} \times \mathscr{M} \to \mathscr{M}$. Namely, $f \circ g$ is to be defined if and only if $d(f) = c(g)$.[3]

The axioms state that the identity morphisms are units, i.e.

$$f \circ \mathrm{id}_{d(f)} = f = \mathrm{id}_{c(f)} \circ f$$

for all morphisms f, and that composition is associative, i.e.

$$f \circ (g \circ h) = (f \circ g) \circ h$$

for any composable morphisms f, g, h — in other words, whenever $d(f) = c(g)$ and $d(g) = c(h)$.

---

[2]To say that the language has two sorts means, semantically, that a model will have two domains of individuals; syntactically, each variable or constant must come with a specified domain or 'type'. For this purpose I write '$x : \mathscr{O}$' to mean that x has type $\mathscr{O}$, and so on. In contrast, I try to reserve '$\in$' to represent the membership relation *in* the theory. It is possible (as in the original presentation (Lawvere, 1964) of ETCS), to formulate category theory in a language with a single sort $\mathscr{M}$ of morphisms — the usual kind of first-order formal language. For my present purposes, using a one-sorted theory would obscure rather than resolve the key issues.

[3]Of course, more formally, we could define $\circ$ in terms of a primitive ternary relation $\alpha(f, g, h)$ of three morphisms, read as 'h is the composite of f and g'. It would be governed by axioms

$$\forall (f, g, h : \mathscr{M}).[\alpha(f, g, h) \to d(f) = c(g)]$$
$$\forall (f, g : \mathscr{M}).[d(f) = c(g) \to \exists!(h : \mathscr{M}).\alpha(f, g, h)].$$

2. Structural Properties and the Object Identity Problem

I take the basic motto of category-theoretic structuralism to be that 'isomorphic objects have all the same properties'. Granted that the properties shared by isomorphic objects are 'structural', the claim then is that all properties are structural. For example, McLarty writes:

> The theory ETCS is structural in the sense that each ETCS set provably has all the same properties as any set isomorphic to it. (McLarty, 2004, p. 48)

Here is a slightly different explanation:

> [Categorical foundations like ETCS] are structuralist in this precise sense: They attribute only structural properties to their objects, that is, only isomorphism-invariant properties. (McLarty, 2005, p. 53)

Note the equivocation between what properties objects have and what properties a theory attributes to them. At a formal level, certainly, 'the properties' McLarty has in mind are just the ones definable in the language of category theory. It is a further metaphysical claim that the definable properties are *the* properties according to some suitably thick notion of 'property'.

I will consider this metaphysical claim in §5. Although I am sympathetic to it, I do not consider it a fundamental part of category-theoretic structuralism. In particular, I leave open for now the more conservative view that the objects of CST are just the Zermelo sets with all their extensional properties. On this conservative view, using category theory is mainly a linguistic maneuver to facilitate the study of structural properties; category-theoretic structuralism is a version of *property structuralism* in the sense of ch. I.2.

In summary, the basic thesis of category theoretic structuralism has two parts: first, that the *definable* properties of objects are isomorphism invariant; second, that the isomorphism-invariant properties are 'structural'. I will focus on the first of these claims, and take the second for granted; but, as I must briefly explain, even the latter requires a pinch of salt. That's because 'isomorphism' is a technical term defined in the language of category theory, and may be interpreted in ways that, on the face of it, have nothing to do with structure. In any category $\mathscr{C}$, an isomorphism is defined to be a morphism with an inverse. That is, f is an isomorphism if and only if

$$\exists(g:\mathscr{M}).[d(g)=c(f)]\wedge[c(g)=d(f)] \\ \wedge[g\circ f=\mathrm{id}_{d(f)}]\wedge[f\circ g=\mathrm{id}_{c(f)}].$$

In standard cases, this is exactly what we want. For example, an isomorphism in the category of groups is a bijective map that suitably relates the two group operations. We can uncontroversially say that isomorphism between groups entails sameness of group-structure. This success, however, depends on the fact that morphisms between groups are *defined* to be structure-preserving according to some antecedent criteria. For the most part, I am happy to focus on such standard cases. With that in mind, let me return to the main claim that all properties definable in category theory are isomorphism-invariant.

Here is McLarty's version of the claim (adapted from (McLarty, 1993)).

THEOREM II.1. *Let $\Psi(X)$ be a sentence in the language of category theory, with a single free variable X of type $\mathscr{O}$. Suppose that o_1, o_2 are isomorphic objects in some category $\mathscr{C}$; then $\Psi(o_1)$ is true if and only $\Psi(o_2)$ is true.*

Here is the proof. We have the usual model-theoretic notion of an isomorphism between categories $\mathscr{C}$ and $\mathscr{D}$.[4] For such an isomorphism F, $\Psi(o_1)$ is true if and only if $\Psi(F(o_1))$ is true. So it suffices to exhibit an automorphism F of $\mathscr{C}$ such that $\Psi(o_1) = o_2$. To do this, choose any isomorphism u from o_1 to o_2. For any object o of $\mathscr{C}$, let $u_o = u$ if $o = o_1$, $u_o = u^{-1}$ if $o = o_2$, and $u_o = i_o$ otherwise. Then here is a suitable F:

(1) $F(o_1) = o_2$ and $F(o_2) = o_1$; otherwise, $F(o) = o$.

(2) For any morphism f, $F(f) = u_{c(f)} \circ f \circ u^{-1}_{d(f)}$.

It is easy to check that this is in fact an automorphism of $\mathscr{C}$.

However, this theorem does not go far enough. The further difficulty is quite simple: the language of two-sorted category theory contains an identity predicate for objects, and this predicate is not isomorphism-invariant. That is, if $A = B$ and $A' \cong A$, it does not follow that $A' = B$.

Call this the *Object Identity Problem*.

3. Three Possible Solutions

3.1. Monadic Predicates? The difficulty just described did not affect the theorem, because the theorem concerned monadic predicates in the pure language of categories, i.e. predicates with a single argument and no constants; it therefore excluded predicates like

$$\mathrm{Eq}(A,B) :\equiv (A = B)$$

(construed either as a two-place predicate of variables A and B or as a monadic predicate of A defined in terms of a constant B). McLarty appears

[4]That is, an isomorphism F is a bijection between the objects of $\mathscr{C}$ and those of $\mathscr{D}$, as well as between the morphisms of $\mathscr{C}$ and those of $\mathscr{D}$, such that (1) $F(\mathrm{id}_o) = \mathrm{id}_{F(o)}$ for every object o of $\mathscr{C}$; (2) $c(F(f)) = F(c(f))$ and $d(F(f)) = F(d(f))$ for every morphism f of $\mathscr{C}$; and (3) $F(f \circ g) = F(f) \circ F(g)$ for all composable morphisms f, g of $\mathscr{C}$.

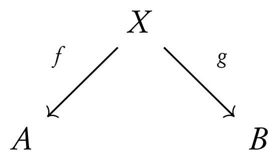
to endorse this as an appropriate restriction on the structuralist thesis. He writes:

> The point is that... objects do "differ" in that each is itself and is not the others; but do not differ in terms of any property stateable without specifying particular objects. (McLarty, 1993, fn. 5)

However, I claim that *even if* we stick to monadic predicates, the theorem does not go far enough. The problem can be framed in different ways depending on the sense in which we understand category theory as a foundation. My point of view in this thesis — because of its direct relevance to HoTT — is that ordinary mathematics is to be developed set-theoretically; what makes the approach 'category-theoretic' is the use of the category of sets (i.e. sets and their functional relations) rather than the cumulative hierarchy of sets (i.e. sets and the membership relation).[5] For example, a group will be a set G with some additional data, that data being specified in the language of the category of sets. But then the definable properties of groups are *not* merely the ones definable in the language of the category of groups, but the ones definable in the language of the category of sets. For example, 'having five elements' is a mathematically interesting, monadic, and indeed isomorphism-invariant, property of groups; but it cannot be defined using only the language of the category of groups. The structuralist motto must be so construed: we want to know that isomorphic *groups* satisfy the same predicates in the language of the category of *sets*. The theorem does not give us what we want.

[5] Awodey (2004) has tried to articulate a very different view of the foundational role of category theory.

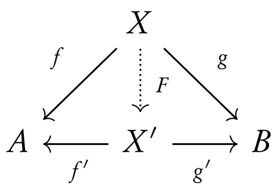
Here is a more detailed example of what can go wrong, simpler than the example of groups, but still — as we will see — mathematically interesting. Start with objects A, B in some category. Let us say that a *span* (over A and B) consists of an object X and two morphisms f, g from X to A and from X to B respectively:



Thus the property of being a span is just

$$\text{isSpan}_{AB}(X, f, g) :\equiv (d(f) = d(g) = X) \wedge (c(f) = A) \wedge (c(g) = B).$$

We can now consider the category of spans. Naturally enough, a morphism of spans from (X, f, g) to (X', f', g') is a morphism F from X to X' such that $f = f' \circ F$ and $g = g' \circ F$:



Such a morphism between spans comes out to be an isomorphism just in case F is an isomorphism between X and X'.

Now, we know from the theorem that isomorphic spans (X', f', g') and (X, f, g) have the same properties expressible in the language of the category of spans. But, using the language of the category of sets, we can define

$$\text{Bad}(X, f, g) :\equiv (X = c(f)). \tag{2}$$

This is a monadic predicate of spans that is not isomorphism-invariant.

EXAMPLE II.2. Let $A = B = X = \{\emptyset\}$, and let $X' = \{\{\emptyset\}\}$. Let $f : X \to A$, $g : X \to B$, $f' : X' \to A$, $g' : X' \to B$ be the unique functions (unique because the codomains are singletons). Then (X, f, g) and (X', f', g') are isomorphic spans, yet $\mathrm{Bad}(X, f, g)$ is true and $\mathrm{Bad}(X', f', g')$ is false.

The problem here is really a form of Benacerraf's objection. Models of the second-order Peano axioms, being isomorphic, *of course* satisfy the same properties expressible in the language of those axioms. The origin of Benacerraf's objection is that those models are *also* Zermelo sets, and the different models differ in their Zermelo-set-theoretic properties, i.e. extensionally. In the present scenario, isomorphic spans *of course* satisfy the same properties expressible in the language of the category of spans (at least to the extent of Theorem II.1). But those spans are *also* constituted by objects and morphisms in the category of sets, and differ with regards to the properties expressible there.

Benacerraf's original objection may appear to be more acute, because we commonly speak of 'the' natural numbers, implying that these are specific mathematical objects (so, seemingly, specific sets). No doubt the ancient pedigree of the natural numbers as objects of mathematical interest does sharpen the point. Still, a look at mathematical practice finds a similar grammatical phenomenon in the case of spans. A span (X, f, g) is said to be a *product* (of A and B) if it satisfies the following property $\mathrm{isProduct}_{AB}(X, f, g)$:

> For any span (X', f', g'), there exists a unique morphism of spans from (X', f', g') to (X, f, g).[6]

[6]Thus, formally:

$$\begin{aligned}\mathrm{isProduct}_{AB}(X,f,g) :\equiv\ & (d(f) = d(g) = X) \wedge (c(f) = A) \wedge (c(g) = B) \wedge \\ & \forall (X' : \mathcal{O}, f', g' : \mathcal{M}).[(d(f') = d(g') = X') \wedge (c(f') = A) \wedge (c(g') = B)] \to \\ & \exists! (F : \mathcal{M}).(d(F) = X') \wedge (c(F) = X) \wedge (f' = f \circ F) \wedge (g' = g \circ F).\end{aligned}$$

It is common among mathematicians to refer to 'the' product of A and B. For example, when it comes to sets, 'the' product $A\times B$ of A and B is said to be the set of ordered pairs (a,b) with $a\in A$ and $b\in B$. What justifies this use of the definite article? Whatever the explanation, this example seems to be on a par with the example of 'the' natural numbers. Just as the Peano axioms determine 'the' natural numbers up to unique isomorphism, but no more, so too the definition of 'product' determines 'the' product up to unique isomorphism, but no more. (As I will argue in §5, 'the set of ordered pairs' is no more uniquely defined than is 'the product'.)

Let me sum up. McLarty implies that category-theoretic structuralism is vindicated by the fact that the monadic predicates of objects define isomorphism-invariant properties. But the predicates for which this is true are the ones in the language of the category of the objects in question. The monadic predicates of these objects in the language of *sets* may not be isomorphism-invariant. So, taking the category of sets as a foundation, the overall theory may attribute to these objects properties like Bad, 'stateable without specifying particular objects' (McLarty), but nonetheless not isomorphism-invariant. It of course remains true that the monadic predicates of *sets* define isomorphism-invariant properties, as in Theorem II.1. But those properties, which depend only on cardinality, are simply not rich enough for the development of mathematics.

3.2. Skeletal Axioms? It may seem that there is a cheap solution to the Object Identity Problem: adopt as an *axiom* that object-identities are structural:

$$\forall(X,Y:\mathcal{O}).(X\cong Y)\rightarrow(X=Y).$$

This is often called a *skeletal* axiom, the image being that every category has a 'skeleton' consisting of one object of each isomorphism class. This trick — if it worked — would be formally cheap, but notably changes the complexion of categorical foundations. For example, it is simply not true that there is a unique set of each isomorphism-class in the cumulative hierarchy. Adopting a skeletal axiom, it would no longer be tenable to consider that the sets in the category of sets just are the sets of the cumulative hierarchy.

However, the skeletal axiom is still not enough. It is true that Bad (2) becomes an isomorphism-invariant property of spans. However, the problem remains for

$$\text{Worse}(X, f, g) :\equiv (g = \text{id}_{c(g)}).$$

EXAMPLE II.3. Let $A = \{\emptyset\}$ and $B = X = X' = \{\emptyset, \{\emptyset\}\}$. Let $f : X \to A$ and $f' : X' \to A$ be the unique maps, $g : X \to B$ the identity map, and $g' : X' \to B$ the unique non-identity map. Then (X, f, g) and (X', f', g') are isomorphic spans, yet $\text{Worse}(X, f, g)$ is true and $\text{Worse}(X', f', g')$ is false.

This shows that the issue is not *just* about object identities. At least some equations between morphisms are also problematic, and this problem is not solved by the skeletal axiom.

This argument sharpens a point made (but then dismissed) by McLarty (2005, fn. 12). He observed that although the skeletal axiom ensures there is only one countably infinite set, there will still be many different systems of natural numbers. These systems of natural numbers differ because their successor functions differ. McLarty discounted this observation because such differences cannot be expressed by a monadic predicate of natural numbers systems. But Worse *is* a monadic predicate of spans.

3.3. Dependent Morphism Types. So far I have explained why two solutions to the Object Identity Problem do not go far enough. The first restricted the structuralist thesis to monadic predicates; the second adopted a skeletal axiom.

The third solution is to remove the identity predicate for objects. However, one cannot do this in a straightforward way. We cannot straightforwardly omit the object-identity predicate from the two-sorted language of categories, because the very axioms of category theory rely on it. For example, the equation

$$c(f) = d(g)$$

uses the object-identity predicate to specify the conditions under which morphisms f and g are composable. Nor can we simply restrict our structuralist thesis to predicates that are 'object-identity free', since (as the example of Worse shows) some object-identity-free predicates are problematic. Even if we could do so, the resulting structuralist thesis would be too weak: important predicates like isProduct contain object-identities.

However, there is a way to distinguish the use of the object-identity predicate in the good cases (the axioms and isProduct), from the way it is used in problematic cases (like Bad). In the former, object identities are only used to *restrict the ranges of morphism-variables.* We can therefore try to restrict ourselves to these good cases by enforcing appropriate syntactic rules. More elegantly, we can obviate the need for object-identities altogether by building such restrictions into the variable types. The result will be a *dependently typed language.* Such languages were originally developed by Makkai (1995) under the name FOLDS (first order logic with dependent sorts).[7]

Here is how it works. To begin in the material mode, we can divide the sort $\mathscr{M}$ of morphisms into subsorts consisting of morphisms with the same domain and codomain:

$$\mathscr{M} = \bigcup_{o,o':\mathscr{O}} \mathscr{M}^{o}_{o'}, \qquad \mathscr{M}^{o}_{o'} := \{f \in \mathscr{M} : [d(f) = o] \wedge [c(f) = o']\}.$$

Now forget about $\mathscr{M}$ itself and take the various $\mathscr{M}^{o}_{o'}$ as primitive: there is to be one sort $\mathscr{O}$ of objects and many sorts $\mathscr{M}^{o}_{o'}$ of morphisms, indexed by pairs of objects. Each morphism-variable is to range over a specified one of these sorts. Crucially, we only allow equations between morphisms, and only between morphisms of the same type. We can still quantify over all morphisms by first quantifying over pairs of objects:

$$\forall(o : \mathscr{O}).\forall(o' : \mathscr{O}).\forall(f : \mathscr{M}^{o}_{o'}) \cdots .$$

To specify the language formally (the complacent reader may skip to the next paragraph!) we begin with a type $\mathscr{O}$. We will have variables but no other terms of type $\mathscr{O}$. For any terms o_1, o_2 of type $\mathscr{O}$, we have a type $\mathscr{M}^{o_1}_{o_2}$. Such a type is said to depend on o_1 and o_2.[8] Given a term o of type $\mathscr{O}$ there is a term $\ulcorner \mathrm{id}_o \urcorner$ of type $\mathscr{M}^{o}_{o}$. Given terms o_1, o_2, o_3 of type $\mathscr{O}$ and terms f, g of types $\mathscr{M}^{o_2}_{o_3}$ and $\mathscr{M}^{o_1}_{o_2}$ respectively, there is a term $\ulcorner f \circ g \urcorner$ of type $\mathscr{M}^{o_1}_{o_3}$. As for sentential formulae, given terms o_1, o_2 of type $\mathscr{O}$ and f, g of type $\mathscr{M}^{o_1}_{o_2}$, we have a wff $\ulcorner f = g \urcorner$. We can combine wffs as usual in propositional logic. The only not-quite-obvious rule is for quantifiers: we must

[7]Makkai had essentially our current purposes in mind; he gives a highly abbreviated version of the current argument at (Makkai, 1995, p. 7).

[8]Note that $\mathscr{M}^{o_1}_{o_2}$, as a linguistic type of terms, is specified by the variables o_1, o_2, *not* by their semantic values. Two such types are the same iff they depend on the same variables.

bind morphism-variables before we bind their domains and codomains. So given a wff F with a free variable f of type $\mathcal{M}^{o_1}_{o_2}$, $\ulcorner\forall(f:\mathcal{M}^{o_1}_{o_2}).F\urcorner$ is a wff. And given a wff F with a free variable o of type $\mathcal{O}$ *and no* free variables of type depending on o, then $\ulcorner\forall(f:\mathcal{M}^{o_1}_{o_2}).F\urcorner$ is a wff. (The same rules go for $\exists$.)

Examples are more useful. In this language, the axiom of identity is

$$\forall(o_1,o_2:\mathcal{O}).\forall(f:\mathcal{M}^{o_1}_{o_2}).(f\circ \mathrm{id}_{o_1}=f=\mathrm{id}_{o_2}\circ f).$$

Similarly, the axiom of associativity is

$$\forall(o_4,o_3,o_2,o_1:\mathcal{O}).\forall(f:\mathcal{M}^{o_3}_{o_4}).\forall(g:\mathcal{M}^{o_2}_{o_3}).\forall(h:\mathcal{M}^{o_1}_{o_2}).$$
$$f\circ(g\circ h)=(f\circ g)\circ h.$$

Here again is the key point: in these axioms, the only use of the identity predicate is between morphisms of the same type. In particular, there is no use for an identity predicate on $\mathcal{O}$. The language also suffices to express the proposition that X,f,g are a product of A and B. As before, I will treat A and B as parameters; the arguments X,f,g have types $\mathcal{O},\mathcal{M}^X_A,\mathcal{M}^X_B$ respectively. Note that being a span is a *trivial* condition on X,f,g; or, better, to say that X,f,g are a span is just to declare the types of these terms. With this in mind, $\mathrm{isProduct}_{AB}(X,f,g)$ has the concise form

$$\mathrm{isProduct}_{AB}(X,f,g):\equiv\forall(X':\mathcal{O},f':\mathcal{M}^{X'}_A,g':\mathcal{M}^{X'}_B).$$
$$\exists!(F:\mathcal{M}^{X'}_X).(f'=f\circ F)\wedge(g'=g\circ F). \tag{3}$$

(Note that the 'unique existence' quantifier $\exists!$ is defined using an identity predicate — but again only between morphisms of the same type $\mathcal{M}^{X'}_X$.)

In contrast, the problematic predicate $\mathrm{Bad}(X,f,g)$ irreducibly involves object-identities. In the two-sorted language, I defined it by the formula

$\ulcorner X = c(f)\urcorner$, or equivalently by $\ulcorner X = A\urcorner$. In the dependently-sorted language, we no longer have a 'codomain' function c, but the second version still makes sense — or *would* make sense if we had an object-identity predicate. We do not.

What about the other problematic predicate Worse(X, f, g) given by $\ulcorner g = \mathrm{id}_B\urcorner$? This predicate does not contain an object-identity predicate. Here is how the current typing system solves the problem. Since (X, f, g) is to denote a span, the variable g must have type $\mathcal{M}_B^X$, while id_B has type $\mathcal{M}_B^B$. Thus the formula $\ulcorner g = \mathrm{id}_B\urcorner$ is not well-formed (cf. fn. 8.)

3.4. Conclusion. The upshot of this discussion is that category theory is best formulated in a dependently typed language. In doing so, we can ensure that

(1) The only identity predicates relate morphisms of the same type.
(2) All properties of interest, like $\mathrm{isProduct}_{AB}$, are definable.
(3) All definable properties and relations are isomorphism-invariant.

Let me expand upon that last point, in case the general statement is not clear. A predicate Ψ in the dependently typed language of categories will, in general, take as arguments some objects $A_i : \mathcal{O}$ and some morphisms $f_{jk}^n : \mathcal{M}_{A_k}^{A_j}$ (with i, j, k, n taking on finitely many values). Suppose we are given some objects $B_i : \mathcal{O}$ and isomorphisms $F_i : \mathcal{M}_{B_i}^{A_i}$. Then the claim is that

$$\Psi(\ldots, A_i, \ldots, f_{jk}^n, \ldots) \text{ is true iff } \Psi(\ldots, B_i, \ldots, F_k \circ f_{jk}^n \circ F_j^{-1}, \ldots) \text{ is true.}$$

4. The Category of Sets

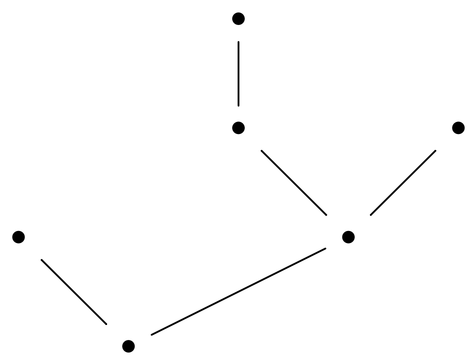
As advertised, the basic kind of 'categorical foundations' I will consider is the kind that develops mathematics set-theoretically – that is, in the language of the category of sets. In this section, I give a partial development of the theory of that category. My goal is, first, to make clear the eventual connection between this theory and HoTT; and second, to address some objections that have been raised to using categorical set theory as a foundation. I discuss those objections in section 5.

4.1. The Structure of Sets. Let me address a preliminary concern. Remember that by 'the category of sets' I mean the category with sets as objects and functions as morphisms. Recall that, as a fallback position, these sets could be the same things studied in Zermelo set theory. Having defined the category of sets in this way, it follows that an isomorphism between sets is a bijection, and therefore that equinumerous sets have the same structure.

But is that *really* true of Zermelo sets? One might think that the structure of a Zermelo set actually determines its extension. Put it this way: if ZST is a piece of mathematics, and mathematics attributes to its objects only structural properties, then the extension of a Zermelo set had better be structural. This becomes even more plausible if we picture Zermelo sets as trees.[9] The children of the root node correspond to the members of the set. The children of *those* nodes correspond to *their* members, and so

[9] A tree is a combinatorial system consisting of 'nodes'. There is a 'root' node, and each node has some other nodes as 'children' (or 'branches'). The exact details do not matter here.

on. For example, the set $\{\emptyset, \{\{\emptyset\}, \emptyset\}\}$ corresponds to a tree like this:



The natural way to define a morphism between trees is as a function between nodes, mapping the root to the root, and preserving the relation of parent to child. Two sets come out isomorphic *as trees* if and only if the sets are extensionally equal. There really is no bar to thinking of the extension of a Zermelo set as part of its structure.

Why then do we claim that *equinumerous* sets have the same structure? In the face of this question, we might try to throw away the crutch of Zermelo set theory and announce that CST reflects some totally different conception of sets; *these* sets don't have the kind of tree structure Zermelo sets do. I will consider such a move in section 5.

But I think the right answer is simpler and more pragmatic. It may be befuddling to ask what the structure of a set is — is it the tree structure, or merely the cardinality? — but everyone agrees what the structure of (say) a *group* is. The structure of a group does not include the tree structure of the underlying set; nor does the structure of the natural numbers. When we are use set theory *as a foundation* for studying these structures, whatever tree structure there may be really is irrelevant. When we do Zermelo set theory *as mathematics*, the tree structure becomes the very thing of interest. And that's just fine: we can study trees in CST.

4.2. The Theory of Sets. Now let us look at what categorical set theory is actually like. A number of different but equivalent axiomatisations are possible (and, too, a variety of extensions have been proposed). The choice of axioms can be guided by different criteria. One criterion is that the axioms should be more or less independent, not only in the logical sense, but in the conceptual sense that the generalisation obtained by omitting one or more of the axioms remains mathematically interesting. This criterion, undoubtedly valuable for some purposes, misleads for others. In the case of the category of sets, the axioms are typically formulated in a way that generalises to the theory of *toposes* (e.g. Mac Lane and Moerdijk, 1997); thus the axioms describe a 'well-pointed W-topos with axiom of choice'. The term 'well-pointed' here means roughly that a morphism can be understood as a relation between elements of its domain and elements of its codomain — as one would expect when a 'morphism' is a function between sets. This axiom, once given, allows for the rest of the theory to be developed in terms of elements and functions as mappings between elements. However, this is not the usual route. Rather, one usually defines what a topos is, in general, and then adduces well-pointedness almost in passing. And this gives rise to the impression that one is supposed to grasp, understand, or justify the axioms independently of the thought that sets have elements and that functions are mappings between elements. As I will later explain (§5), this has been considered to weigh against categorical set theory on epistemic grounds.

Thus my strategy in this section is to give axioms that are stated primarily in terms of elements. This puts categorical set theory on firmer epistemic ground than is generally recognised. It may, for some tastes, not be enough; afterwards I will discuss some of the remaining epistemic quandries. It must also be admitted that there is still something formally awkward about the way elements are treated in CST. This is arguably a point on which HoTT will offer much improvement.

The axioms can be divided into two groups. The axioms under heading *I* below below ensure that elements and functions behave as expected; those under heading *II* postulate the existence of certain sets. More precisely, they postulate the existence of sets satisfying certain properties that characterise them up to isomorphism. In a sense, then, they only postulate the existence of certain cardinalities; they have the flavour of axioms of infinity.

I. Elements and Functions

The first step is to explain why the language of the category of sets is rich enough to speak about the elements of sets. (It must be done. But the reader convinced that it *can* be done might well skip to heading *II*.)

1. SINGLETONS AND ELEMENTS.

First we characterise the singleton sets.

DEFINITION II.4. An object A is *singleton* (or, in the usual categorical terminology, *terminal*), if and only if it receives a *unique* function from every object:

$$\forall(X : \mathscr{O}).\exists!(f : \mathscr{M}_A^X).$$

This brings us to the first axiom:

AXIOM II.5. *There exists a singleton set.*

Now suppose that we fix upon a singleton set — call it **1**. We expect that for any set S, the functions $\mathbf{1} \to S$ exactly correspond to the elements of S. We have a type $\mathscr{M}^1_S$ of such morphisms; so this type parameterises elements of S. It is convenient to refer to the elements of this type simply *as* elements of S. But it must be bourne in mind that if $\mathbf{1}'$ is another singleton set, then $\mathscr{M}^{\mathbf{1}'}_S$ has equal claim to be 'the' type of elements of S.

This is not supposed to mean that S has elements *relative to* the choice of a singleton set. There are a number of ways to explain what it *is* supposed to mean; but this is the basic awkwardness of categorical set theory. Perhaps the most obvious move would be to postulate, for each object S, a sort $\mathscr{E}_S$ of 'elements of S', and to add axioms that allow one to pass from $\mathscr{M}^1_S$ to $\mathscr{E}_S$, for any terminal object $\mathbf{1}$. To look ahead, this is essentially what is done in HoTT. But, by introducing the element sorts $\mathscr{E}_S$, it takes us beyond the pure language of categories I am considering here.

Still, the reason that we can at least make do with the pure language of categories is that, if we fix upon a singleton set $\mathbf{1}$, we can quantify over elements of S. When convenient, then, I will write

$$s \in S$$

to mean that s is an element of S (i.e. a morphism of type $\mathscr{M}^1_S$). The particular choice of a singleton set is 'harmless' in the sense that any singletons $\mathbf{1}$ and $\mathbf{1}'$ have all the same definable properties (since, as follows directly from the definition, they are isomorphic, and indeed uniquely so). If you and I happen to choose different singleton sets, there is a unique way to translate between us. I will come back to this issue in Chapter III. For now let us simply suppose that $\mathbf{1}$ is some particular singleton set.

2. Function Extensionality and Plentitude.

Moving on, then, we expect that a morphism should determine a mapping between elements. And that is just what happens. Given $f : \mathscr{M}^S_T$ and

$s \in S$, we obtain $f(s) :\equiv f \circ s \in T$. We can then ask whether f is *injective* ($f(s) = f(s') \to s = s'$); we can ask whether a given $t \in T$ lies in the image of f ($\exists(s \in S).f(s) = t$); and if every t lies in the image of f, we can say that f is *surjective*. So we are off and running.

However, we don't just expect that morphisms should *determine* mappings. They should be individuated the same way mappings are: by their values. This is essentially the key 'well-pointedness' axiom, which I will call by the more intuitive name 'Function Extensionality'.

AXIOM II.6 (Function Extensionality). *Morphisms* $f, g : \mathcal{M}_T^S$ *are equal if and only if* $f(p) = g(p)$ *for all* $p \in S$.

This allows us to understand a function as a relation or correspondence between elements in the usual way. The next axiom allows functions to be defined by formulae. In standard treatments it is a theorem instead of an axiom — note how it, like Function Extensionality, uses the notion of 'element' in a fundamental way.

AXIOM II.7 (Function Plenitude). *Suppose* Ψ *is a binary predicate on sets* A *and* B.[10] *Then*

$$[\forall(a \in A).\exists(b \in B).\Psi(a,b)] \to [\exists(f : \mathcal{M}_B^A).\forall(a \in A).\Psi(a, f(a))].$$

In other words, if Ψ determines at least one b for every a, then there is a function f mapping as to bs. Two points to note. First, if Ψ determines a unique b for every a — so that b 'depends functionally' on a — then Function Extensionality ensures that the f is unique. Thus one may genuinely define functions by formulae. Second, Function Plentitude entails

[10]In other words, it is a sentence with two free variables of type $\mathcal{M}_A^1$ and $\mathcal{M}_B^1$.

a standard version of the axiom of choice: every surjective function has a right-inverse.

II. Existence Axioms

Now that I have explained how to talk about functions and elements in the category of sets, let us look at its basic ontology.

3. Cardinality 0 and Cardinality 1

Axiom II.8. *There exists a set* $\mathsf{0}$ *with no elements, and [to repeat:] a set* $\mathbf{1}$ *with exactly one element.*

4. Cardinal Addition

Suppose that sets A_1 and A_2 have cardinalities $\#A_1$ and $\#A_2$ respectively; we expect there to be a set of cardinality $\#A_1 + \#A_2$. This means we can partition the set into two parts, one of which is isomorphic with A_1 and the other of which is isomorphic with A_2.

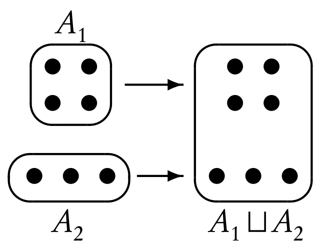


Figure 1. A coproduct $A_1 \sqcup A_2$.

Definition II.9. A set $A_1 \sqcup A_2$ is the *coproduct* (or disjoint union) of A_1 and A_2 with respect to functions $\mathrm{in}_1 : \mathscr{M}^{A_1}_{A_1 \sqcup A_2}$ and $\mathrm{in}_1 : \mathscr{M}^{A_2}_{A_1 \sqcup A_2}$ if every element of $A_1 \sqcup A_2$ is uniquely of the form $\mathrm{in}_1(a_1)$ or $\mathrm{in}_2(a_2)$.

(These 'in_1' and 'in_2' are the first and second *inclusions*.)

AXIOM II.10. *For any sets A_1 and A_2 there is a coproduct $A_1 \sqcup A_2$.*

5. CARDINAL MULTIPLICATION.

For any sets A_1, A_2 we expect there to be a set of cardinality $\#A_1 \times \#A_2$. What this means is that it is the right size to parameterise ordered pairs (a, b).

DEFINITION II.11. A set $A_1 \times A_2$ is a *product* of A_1 and A_2 with respect to $\mathrm{pr}_1 : \mathscr{M}_{A_1}^{A_1 \times A_2}$ and $\mathrm{pr}_2 : \mathscr{M}_{A_2}^{A_1 \times A_2}$ if for each $a_1 \in A_1$ and $a_2 \in A_2$ there is a unique $(a_1, a_2) \in A_1 \times A_2$ such that $\mathrm{pr}_1(a_1, a_2) = a_1$ and $\mathrm{pr}_2(a_1, a_2) = a_2$.

(These pr_1 and pr_2 are the first and second *projections*.)

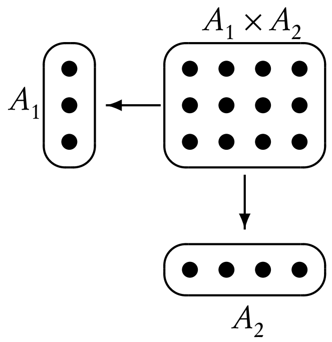


FIGURE 2. A product $A_1 \times A_2$.

AXIOM II.12. *For any A_1 and A_2 there is a product $A_1 \times A_2$.*

The notion of a product was already introduced (3) for objects in an arbitrary category. The characterisation of 'the product of two sets' I have used here only really works in the category of sets, since it talks about elements of objects. But it is not hard to check that any $A \times B, \mathrm{pr}_1, \mathrm{pr}_2$ satisfying Definition II.11 is indeed a product in the general sense, using Function Extensionality and Function Plenitude.

6. CARDINAL EXPONENTIATION.

For any sets A, B, we expect there to be a set of cardinality $\#B^{\#A}$ — that is, of the right size to parameterise functions from A to B.

DEFINITION II.13. For any $A, B : \mathcal{O}$, B^A is a *mapping set* with respect to an 'evaluation' morphism $\mathrm{eval}\colon B^A \times A \to B$ if

$$\forall (f : \mathcal{M}_B^A).\exists!(\alpha \in B^A).\forall (a \in A).\, \mathrm{eval}(\alpha, a) = f(a).$$

AXIOM II.14. *For any $A, B : \mathcal{O}$, there is a mapping set B^A,* eval.

Note that, strictly speaking, the mapping set is also relative to the choice of a particular product $B^A \times A$. Just as all the singleton sets were isomorphic, and so indiscernible, so too are all the products isomorphic and indiscernible. But it is still an awkward situation. I return to this issue in III.1.2.

7. COMPREHENSION.

Given a set A and a predicate Ψ on A, we expect there to be a set with the same cardinality as the Ψs. Although I will call this an 'extension' of Ψ, it needn't somehow *consist of* the Ψs; it just has to map injectively onto them.

DEFINITION II.15. Suppose that Ψ is a predicate on A. Then a set E_Ψ is an *extension* of Ψ with respect to $u : \mathcal{M}_A^{E_\Psi}$ if and only if u is injective and

$$\forall (a \in A).[\Psi(a) \leftrightarrow \exists (x : E_\Psi).u(x) = a].$$

AXIOM II.16 (Comprehension). *Suppose Ψ is a predicate on a set A. Then Ψ has an extension.*

8. INFINITY.

We expect there to be a countably infinite set.

DEFINITION II.17. A set $\mathbb{N}$ is a 'natural numbers object' with respect to $0 \in \mathbb{N}$ and succ : $\mathscr{M}_{\mathbb{N}}^{\mathbb{N}}$ if it satisfies the second-order Peano axioms. (In the following form: succ is injective; 0 is not in the image of succ; and for any set K and $f : \mathscr{M}_{\mathbb{N}}^{K}$, if 0 is in the image of f and succ(n) is in the image of f whenever n is, then f is surjective.)

AXIOM II.18 (Infinity). *There exists a natural numbers object.*

POWER SETS AND OTHER OBJECTS. Standard axiom sets postulate the existence of a 'power set' $P(A)$ for each set A, or else the existence of a 'sub-object classifier' — roughly, a set of truth values. The two are related in that if T is a sub-object classifier, then T^A is a power-set for A. We do not need an additional axiom for this; any two-element set will do for T. For example, $\mathbf{1} \sqcup \mathbf{1}$ will do. The use for such an object will reappear dramatically in ch. III.3.

5. Is Categorical Set Theory Autonomous?

It is time to revisit the relationship between CST and Zermelo set theory. Again, my fallback position is that the objects of categorical set theory just are Zermelo sets. What makes categorical set theory special is that it adopts a language in which only structural properties can be expressed. On this view, sets have non-structural properties, but they are not mathematically interesting. (Again, I don't mean by this that ZST isn't mathematically interesting. I mean that the non-structural properties of sets are not relevant for the use to which sets are typically put in mathematics.)

Thus the fallback position is that categorical set theory embodies *property structuralism*.

Objections can be raised both to property structuralism in general and to this version of property structuralism. I will consider some of the each kind in Chapter III. First I want to examine the prospects for a more ambitious reading of categorical set theory. This more ambitious reading would make three claims. First, it would deny that the objects of CST are Zermelo sets. Second, it would deny that our grasp of these objects and their properties depends on our grasp of Zermelo set theory. Third, it would claim that CST articulates a version of *object structuralism* (cf. I.2).

Let me begin with the third claim. The idea would be that, not only are all the properties definable in CST structural, but that these definable properties are the only properties its objects have. As McLarty says, when we speak the language of CST, we are not 'ignoring' anything about the objects in question (McLarty, 1993, p. 496). Here are two reasons for scepticism. First, the version of CST I have arrived at does not have an identity predicate for sets. This is not particularly troubling if we are just formally excluding equations from our language. But the object-structuralist seems to claim that there is just no question of whether sets A and B are equal. That is strange. The second feature of CST that should trouble the object-structuralist is the complicated system of types. The metaphysical import of type distinctions has never really been clear, but the situation here is particularly egregious. If A is a set and B is a set, and a is an element of A and b an element of B, then there seems to be no question of whether a equals b. One might have thought, 'Well, at least if A equals B we can ask whether a equals b.' But there is no question of whether A equals B! It is hard to take this seriously as more than a linguistic gambit to rule out non-structural predicates. (Getting around these problems is the key to the object-structuralist reading of homotopy type theory; that is the subject of Chapter IV.)

Note that this line of criticism tells less strongly against the idea that the *sets* of CST have only structural properties. The big problem came when we considered more complicated systems *constructed* from sets, like spans. Only then did we abjure object-identities and embrace dependent typing. The two-sorted version of CST is metaphysically less mysterious, and it assigns to sets themselves only structural properties, at least to the extent of Theorem II.1. It is possible to take seriously the idea that this basic version of CST articulates a sui generis (and *perhaps* object-structuralist) conception of sets.

That leads me back to the first two 'ambitious' claims. I will frame the discussion using some objections raised by Linnebo and Pettigrew (2011). They question the 'conceptual' and 'justificatory' autonomy of CST from Zermelo set theory. They summarise the general issue like this.

> [A] putative foundation for mathematics must boast more than mere logical autonomy with respect to [Zermelo] set theory if it is to be truly autonomous. It must be possible not only to formulate the foundation without presupposing a theory of sets; it must be possible also to *understand* it and to *justify its claims* without such a presupposition. (Linnebo and Pettigrew, 2011, p. 241)

I will consider whether categorical set theory, as I presented it, can meet these conditions of autonomy. Before doing so, it is worth emphasising again that this kind of autonomy is not really necessary for most 'foundational' purposes. First, from the point of view of property structuralism, category theory provides a way of restricting attention to the structural properties of set-theoretic systems; moreover, experience has shown it to be an effective way of studying those properties. None of that would change if the underlying sets were Zermelo sets and had to be understood as such. Second, even if we *do* deny that the sets of CST are Zermelo sets, questions about their nature are somewhat independent of questions about how we grasp their properties. Perhaps the axioms of CST really do point to some distinctive entities; perhaps they are cardinality structures in the sense of Shapiro, or else sets of *lauter Einsen* in the sense of Cantor.[11] Still, it might be that we grasp the properties of these objects by abstracting from Zermelo sets, and that we must point to Zermelo sets (or, just as likely, sets of concreta) when justifying our foundational axioms. As an analogy, we may well understand and justify arithmetic by counting beans; it doesn't follow that natural numbers *are* beans or that the Peano axioms are foundationally suspect.

The Linnebo-Pettigrew objection can thus only clarify and limit, rather than overturn, the foundational project of categorical set theory. With that in mind, let me turn to the specific objections raised against (first) the conceptual autonomy, and (more briefly) the justificatory autonomy of categorical set theory. On conceptual autonomy, Linnebo and Pettigrew consider two particular objections. The first (which they attribute to Dan Isaacson) is that

[11]As it is frequently suggested: see e.g. (Lawvere, 1994) and (Linnebo and Pettigrew, 2011).

> whenever we come to explain these axioms to those unfamiliar with them, we inevitably appeal to the membership relation, the subset relation, the notion of ordered pair, the notion of a function as a set of ordered pairs, and so on. That is, at the point of explanation, the mapping-theoretic presentation is abandoned in favour of a more orthodox presentation, which is required to allow us to understand the axioms. Thus, ETCS does not have conceptual autonomy. (p. 242)

Part of the response to this objection must be that the sets of CST do have members, and we are allowed any explanation that relies on that fact. The question, when it comes to membership, is whether *membership between sets* is necessary for an understanding of the axioms. And it does not seem to be.

For example, it is perfectly clear that the product axiom is motivated by 'the notion of ordered pair'. That is, we can try to understand the product of A and B as the set of ordered pairs (a, b). On first glance, this appears to be an extensional definition, the sort of thing that makes sense in Zermelo set theory but not in CST. But that is not true. It is not an extensional definition until we say what '(a, b)' means. In Zermelo set theory, it must denote a set. And Benacerraf's objection applies to ordered pairs just as much as it applies to natural numbers: there are many ways to *encode* an ordered pair as a set, but seemingly no set that the ordered pair *is*. For example, the ordered pair (a, b) of any two sets a and b is typically said to be $\{\{a\}, \{a, b\}\}$, following Kuratowski. But there is something fundamentally arbitrary about this expression. It is no part of any prior conception of 'ordered pair', for example, that $\{a\} \in (a, b)$, even though this holds on Kuratowski's definition. Nor does it seem to be something that we might discover to be true. After all, what makes it the case that (a,b) is $\{\{a\},\{a,b\}\}$ rather than $\{\{a,b\},\{b\}\}$, or indeed (following Wiener) $\{\{\{a\},\emptyset\},\{\{b\}\}\}$? With the latter two definitions, it would not be true that $\{a\} \in (a,b)$. Kuratowski's pairing is, pretty obviously, a matter of convention.

Moreover, how does one verify that Kuratowski's definition is an adequate one? First, one observes that any $a \in A$ and $b \in B$ determine a unique Kuratowski pair. Then one checks conversely that any Kuratowski pair comes from a unique $a \in A$ and $b \in B$. But this is just another way of saying that the set $(A \times B)_K$ of Kuratowski pairs comes with functions $\mathrm{pr}_1 : (A \times B)_K \to A$ and $\mathrm{pr}_2 : (A \times B)_K \to B$ such that for each $a \in A$ and $b \in B$ there is a unique Kuratowski pair x such that $\mathrm{pr}_1(x) = a$ and $\mathrm{pr}_2(x) = b$. The existence of appropriate functions pr_1 and pr_2 is *the* criterion for any definition of 'ordered pair' in Zermelo set theory to be formally adequate. To be precise, it isn't the mere *existence* of pr_1 and pr_2 that matters. Consider again the question of $\{\{a\},\{a,b\}\}$ versus $\{\{a,b\},\{b\}\}$. According to the Kuratowski choice of pr_1 and pr_2, the first of these is (a,b) and the second is (b,a). It is not the *extension* of the set $(A \times B)_K$ of Kuratowski pairs that makes it the product of A and B as opposed to the product of B and A. It is only the product *vis a vis* particular choices of pr_1 and pr_2.

In short, we understand that Kuratowski's definition provides a reasonable analysis of ordered pairs just because it satisfies the definition of a product used in categorical set theory. The notion of ordered pair is something conceptually independent of Zermelo set theory; if anything, the notion of a product set used in CST must be used to explain the use of Kuratowski pairs.

Essentially the same point can be made about appeals to 'the notion of a function as a set of ordered pairs'. In the first case, I do not know that anyone conceives of functions *as* sets of ordered pairs; we constantly distinguish between functions and their graphs. Second, even if a function were a set of ordered pairs, it would not follow that a function is a set of *Kuratowski* pairs. That is, the notion of 'ordered pair' that allows us to conceptualise graphs of functions is certainly prior to the way in which ordered pairs and functions are encoded in Zermelo set theory. The encoding of functions as sets of ordered pairs (of any sort) has a conventional aspect to it. As Lawvere (1964, p. 12) mentions, functions could be represented equally well by their 'cographs' rather than their graphs.[12] There is nothing that makes the set of graphs, *rather than* the set of cographs, the set of functions between A and B. Both count as mapping sets because they stand in appropriate relations to A and B. Categorical set theory precisely spells out what relations are required (Definition II.13). If we admit that it is legitimate to think of the set of graphs as a mapping set, then we should admit that what makes it legitimate are the functional relations described in categorical set theory.

There is, perhaps, one case in which ZST might seem to have conceptual priority: the case of power sets. The power set of X has a very natural extensional definition as 'the set of all subsets' of X. However, the fact is that this extensional definition does no work in ordinary mathematics.

[12]The cograph of a function $f : X \to Y$ is a set of subsets of the disjoint union $X \sqcup Y$. A subset of $X \sqcup Y$ belongs to the cograph of f if and only if it contains a single element y of Y and the preimage of y under f.

Instead of the set of all subsets of X, one could consider the set of all characteristic functions on X. The set of subsets and the set of characteristic functions both stand in the relevant functional relations to X. Both can play the role of 'power set'. Based on what I have said, the notion of a set of characteristic functions is conceptually independent of Zermelo set theory.

The second objection considered by Linnebo and Pettigrew (which they attribute to John Mayberry) is also relevant here.

> The objector submits that the only precise account of the notion of mapping that captures the level of idealization that is required in modern mathematics is given by the definition of a function as a set of ordered pairs that represents a many-one or one-one relation, and this definition belongs essentially to orthodox set theory.

The response of Linnebo and Pettigrew is adequate here: CST theorises about functions axiomatically rather than by reducing them to something else. To that I can add that the representation of a function by a set of ordered pairs is entirely available in CST: each graph defines a function, each function a graph. But CST gives the correct account of why this happens: the set of graphs 'is' a set of functions because it satisfies Definition II.13.

In summary, I think that the case for the *conceptual* autonomy of categorical set theory is very strong. At a minimum, the traffic goes both way. The concepts of categorical set theory explain why it is we use Zermelo set theory the way we do. But the question of *justification* is more difficult. What the use of Kuratowski pairs does in ZST is establish the *existence* of a product set. It does so on the basis of general principles of comprehension (cf. Boolos, 1971). In contrast, CST simply postulates that products exist. There may be a naturalistic justification for such existence claims, its being undeniable that mathematicians systematically make use of product sets. But, as Linnebo and Pettigrew rightly point out, if one wants a detailed account of such existence claims even within mathematical practice, it is hard to see past Kuratowski pairs. Similar worries apply to sets of functions. Encoding functions by their graphs, and forming the set of all such graphs, gives a detailed justification for the exponential axiom.

However, the situation is not so clear-cut. It is bound to be controversial what counts as a justification and how much detail is required. True, the iterative conception of sets yields a general story about existence claims; but what justifies the iterative conception? One might contend that the particular existence axioms of CST are just as clear.

Here are two brief stories along those lines. I do not claim that either of them is completely satisfactory; the question, in relation to Zermelo set theory, is whether they are more or less mysterious than the distinction between several objects on the one hand and the single set that 'collects' them on the other. The first story was urged on my by David Wallace. On this story, the categorical conception of sets is that sets are the domains and codomains of functions. If we know what functions are, then we know what binary functions are; there must exist product-sets as their domains. Mapping sets are the domains of higher-order functions, and, perhaps, so on. The second story hews more closely to the version of categorical set theory I presented in this chapter. This story emphasises that the existence axioms are pure cardinality claims. We take cardinality as the basic concept. Once we have fixed on A and B, anything of the right cardinality will do as $A \times B$. The cardinality of a disjoint union $A \sqcup B$ is the sum of the cardinalities of A and B; the cardinality of $A \times B$ is their product; and the cardinality of A^B is the exponential. The main existence axioms of CST assert the possibility of cardinal arithmetic. To be sure, this is not a detailed justification for such axioms. The iterative conception of sets gives a longer story about how to construct such cardinals. But it would suffice to get them any way we can.[13]

6. Conclusion

In conclusion, categorical set theory is best understood as a property-structuralist, rather than object-structuralist, foundation. To fulfill even this property structuralism, we must resort to dependent typing. On my fallback view, the sets of CST just are Zermelo sets; but there is also scope for a more ambitious view according to which the sets of CST are sui generis, and according to which their monadic properties are structural. Moreover, we can grasp, and *perhaps* even justify, the axioms of CST without reference to Zermelo set theory.

[13]One alternative to ZST in this role might be plural quantification and mereology, as suggested in (Hellman, 2006).

CHAPTER III

Martin-Löf's Dependent Type Theory

In the previous chapter, I explained how and why one might develop categorical set theory in a first order language with dependent sorts (Makkai's FOLDS). It would be interesting to examine Makkai's project in its own right, but homotopy type theory leads us down another path.

Homotopy type theory differs from categorical set theory in FOLDS in four main ways. Three of them are inherited from Martin-Löf's dependent type theory (MLTT), on which HoTT builds. First, the existence axioms are explicitly constructive. For example, HoTT has a name $\ulcorner \mathbf{1} \urcorner$ for a terminal object, whereas CST, at least in my presentation of it, merely posited that such an object does exist. Second, HoTT uses a system of universes, instead of a single domain of sets. Besides extending the ontology, this effectively makes HoTT into a higher-order theory.[1] Third, instead of building on a background logic like FOLDS, HoTT *includes* logic by introducing a type of propositions (roughly: truth values) on a par with the other objects of the theory.

In this chapter, I investigate these three features of HoTT. Perhaps the most important argument for the big picture is contained in section 4. There I reject an intuitionistic reading of dependent type theory.

[1] Makkai saw FOLDS as a first-order fragment of Martin-Löf type theory (which preceded it). That may be formally right, but I think it more informative to say that FOLDS is a relative of first-order logic, while Martin-Löf's theory is a relative of set theory (or of other traditional type theories).

I leave to subsequent chapters an investigation of the fourth important feature, which distinguishes HoTT from the basic version of dependent type theory: its treatment of identity between sets.

1. Two More Problems for Categorical Set Theory

In this section, I consider criticisms of categorical set theory that motivate two of the novel features of dependent type theory: the use of universes, and the use of constructive existence axioms.

1.1. Hellman's Open-Endedness Objection. One of the key discussions of category-theoretic structuralism is contained in the exchange between Hellman (2003), Awodey (2004), and McLarty (2004, 2005). Many of Helmann's objections traded on the idea that the axioms of category theory are to be understood schematically, much like those of group theory. He argued that, as such, they could not play a foundational role. Awodey effectively denied the inference, while McLarty responded that, on his understanding of the position, the foundation is not category theory per se, but theories of specific categories, like the category of sets. McLarty's is the point of view adopted here: CST plays the foundational role, and this blocks most of Hellman's original objections.

One objection remains. As Hellman later wrote

> What remains problematic... regarding [CST] is its apparent commitment to a fixed, presumably maximal, real-world universe of sets, "the category of sets". This just strikes me as a convenient fiction. (Hellman, 2006, p. 6)

It is important to note that his concerns apply to Zermelo set theories just as much as to CST. He is more specifically worried about two features of the situation. The first is the plurality of set theories:

> First, there is the question of multiplicity of conceptions of sets, e.g. non-well-founded as well as well-founded, possibly choice-less as well as with choice.... Presumably, all of these conceptions are mathematically legitimate, and it would be arbitrary to treat just one as ontologically privileged.

I responded to similar concerns in ch. II.4. One way to study sets mathematically is by adopting them foundationally, as the raw material of mathematics. But another is to *represent their structure* using whatever raw materials one has at hand. One can study different conceptions of sets mathematically the same way one can study unicorn populations mathematically. Categorical set theory provides enough raw materials to do so. Its sets are not privileged 'ontologically' but pragmatically: they provide for the perspicacious development of contemporary mathematics and, in particular, for the study of structural properties.

Hellman's second concern is more serious.

> Whatever domain of sets we recognize can be transcended by the very operations that set theory seeks to codify, collecting, collecting everything "already collected".... Set theoretic structuralism can be faulted precisely for failing to apply to set theory itself, especially in regard to the very multiplicity of universes of sets that it naturally engenders. Categorical structuralism promises to do better, but it is hard put to keep that promise if it falls back on a maximal universe of sets or, more generally, on an absolute notion of "large category." (pp. 6–7)

The initial formulation here does not sit well with categorical set theory: it is not obvious that categorical set theory seeks to codify 'collecting'; that is, rather, the aim of Zermelo set theory. But the last sentence of the quotation suggests a particular form of the problem to which the category-theoretic structuralist should feel more sympathy. Not only does one want to consider groups as structured sets, one would also like to consider the category of groups in its own right. But if this is in some sense a collection of structured sets, it is not itself a set. There are actually two difficulties here. First, there is the usual cardinality worry, that the collection of groups is too large to be a set. Second, there is a worry about extensionality: what does it mean to speak of 'the category of groups' if we eschew extensional definitions of collections?

The method of universes addresses the first of these worries (and I will say something about the second in ch. V.1). Instead of a single domain $\mathscr{O}$ of 'all' sets, the theory posits a sequence

$$\mathscr{U} \subset \mathscr{U}' \subset \mathscr{U}'' \subset \cdots$$

of increasingly large universes, each containing its predecessor as an object.[2] Each universe satisfies the axioms of categorical set theory: any two objects in $\mathscr{U}$ have a product in $\mathscr{U}$, and so on. One always quantifies over one of the universes; there is no way to quantify over their union. Thus, while one can't talk about the category of absolutely all groups, one can talk about the category of all groups (implicitly: in the universe $\mathscr{U}$); it is

[2]For a good account of universes in set theory see Feferman (2004). Roughly speaking, the universes are a hierarchy of inaccessible cardinals.

a category whose sort of objects is an element of $\mathscr{U}'$. It is, of course, legitimate to ask whether this method of universes is ultimately satisfactory. But it does overcome a narrow reading of Hellman's objection.[3]

1.2. Existence and Reference. There is an unresolved awkwardness in my presentation of categorical set theory. It had several mere existence axioms, stated (if informally) in the pure language of categories. For example: for any A and B there exists a product set. In contrast, one could introduce into the language a function symbol $\times$ for forming new sets from pairs of old, and take it as an axiom that $A \times B$ is a product set of A and B with respect to certain morphisms. It is undeniable that the latter approach better reflects ordinary mathematical reasoning.

The evidence for this can be found even in my presentation of the theory. To define the elements of a set, I effectively introduced a constant **1** for a singleton set. To define mapping sets, I effectively introduced a particular product $B^A \times A$. And so on. I could have done otherwise, and worked very strictly within the unadulterated language of category theory. For example, when defining the notion of a product, instead of saying that any $a_1 : \mathscr{M}^{\mathbf{1}}_{A_1}$ and any $a_2 : \mathscr{M}^{\mathbf{1}}_{A_2}$ come from a unique $(a_1, a_2) : \mathscr{M}^{\mathbf{1}}_{A_1 \times A_2}$, I could have said that for *any* terminal object **1**, and any $a : \mathscr{M}^{\mathbf{1}}_{A} \ldots$ You could read me as implicitly replacing constants by appropriately bound variables. Doing this would be harmless as far as the Object Identity Problem goes, once we have got rid of object-identity predicates. But doing it *explicitly* would have made my presentation of the theory much different. More importantly, it would be a significant (if tedious) undertaking to recast ordinary

[3]The reader may note a certain tension between my responses to the two long quotations from Hellman. One way to study the category of all groups in $\mathscr{U}$ is to include it in a larger universe $\mathscr{U}'$. But another way is just to *represent* it in the basic universe $\mathscr{U}$. This second strategy is one possible spin on Awodey's 'top-down' categorical structuralism (Awodey, 2004). Still, there may be occasions on which when one *really* needs more raw material.

mathematics in this way. Although apparent reference to (say) 'the product of A and B' can be removed from mathematical discourse by appropriate quantification, this paraphrasing seriously misrepresents ordinary mathematical reasoning (Breckenridge and Magidor, 2012).

As I said, it is *possible* to develop a version of categorical set theory with constants rather than with mere existence axioms. Roughly speaking, Martin-Löf's type theory follows this route. However, such a move requires interpretation. There is a serious tension between the introduction of a function symbol $\times$ into the language, on the one hand, and the view I took on product sets in ch. II.5, on the other. The obvious explanation of talk about 'the product of A and B' is that the term $\ulcorner A \times B \urcorner$ has some specific intended interpretation — the very set of ordered pairs. But, on the contrary, I have argued that 'the set of ordered pairs' does not latch onto any particular thing, at least in Zermelo set theory. The phrase 'ordered pair' does not pick out some particular things that can then be collected in a set; rather, any set potentially counts as a set of ordered pairs vis a vis some functional relations.

Now, in the end, HoTT provides a distinctive way of resolving this tension. On the face of it, at least, A and B *have a unique product* in HoTT. That is, there is only *one* object that stands in the right relations to A and B. So one can legitimately refer to *the* product of A and B. A natural interpretation, inspired by ante rem structuralism, is that $\ulcorner A \times B \urcorner$ refers to the structure had by any product of A and B; the thought is that this structure *itself* counts as a product. That is the picture I investigate in Chapter IV. However, while we are making the transition from categorical set theory to homotopy type theory, it is worth asking whether the use of constants in CST can be defended in its own right. I think it can, and that must qualify the structuralist motivation for homotopy type theory. The defence I have in mind would legitimate the use of $\ulcorner A \times B \urcorner$ as (something like) a singular term, while denying it any specific intended referent.

The idea is that the language of category theory is so peculiar that singular terms may not have quite the semantic roles that we expect. Although it is not true that any two products of A and B are equal, they aren't *not* equal either, as far as the language goes; there is no identity predicate to relate them. Since there is no question in the language of whether $A \times B$ is one product versus 'another', it is not obvious why the semantic role of $A \times B$ must be to refer to one product versus another — even if such distinctions can be drawn in the metalanguage. A bit more concretely: the 'semantic role' of $\ulcorner A \times B \urcorner$ is something like the way in which $\ulcorner A \times B \urcorner$ contributes to the truth conditions of sentences containing it. For example, suppose that $\ulcorner \Psi(\mathrm{JC}) \urcorner$ is true if and only if Ψ is true of Caesar; this determines Caesar as the referent of 'JC' precisely because Caesar is discernible using various predicates Ψ. The semantic role of 'JC' is also 'complete' in the sense that nothing must be added to determine the truth values of sentences containing it. In contrast, the semantic role of a variable x is 'incomplete': sentences $\ulcorner \Psi(x) \urcorner$ only have truth values relative to an assignment, or, better, when the sentence is completed by quantification.

Here is the view that I suggest.[4] The semantic role of $\ulcorner A \times B \urcorner$ is like that of 'JC', and unlike that of x, in that it is complete. Every sentence

[4]The more general views of (Shapiro, 2012) and (Breckenridge and Magidor, 2012) similarly trade on novel semantics. My view seeks to exploit a very special feature of the situation, namely the linguistic indiscernibility of objects.

$\ulcorner\Psi(A \times B)\urcorner$ has a truth value, and there is no need to paraphrase or supplement such a sentence with quantifiers. But $\ulcorner A \times B\urcorner$ is nonetheless *unlike* 'JC' in that no unique object is appropriately coordinated with those truth values. Of course, certain objects — the products — are relevantly coordinated with the truth values; but not one product instead of another. The point, of course, is that the products are indiscernible in the formal language; one might say that $\ulcorner A \times B\urcorner$ *refers undiscerningly* to the products of A and B. The semantic role of $\ulcorner A \times B\urcorner$ licenses all the usual reasoning about 'the' product of A and B; but we should not necessarily conclude that this semantic role is like that of 'JC' in all respects.

I warned that the sketched account would tend to undermine the motivation for object structuralism and the move to homotopy type theory. The reason should now be clear. Part of the motivation for object structuralism is to provide a specific intended referent for (say) 'the natural numbers'. But perhaps we can explain the definite article by other means. On the first account I considered, apparent reference to the natural numbers is to be eliminated in favour of universal quantification over natural numbers systems (*pace* McLarty, 1993, fn. 7). But this misrepresents ordinary mathematical reasoning. On my second, preferred account, it is entirely legitimate to speak of 'the natural numbers'. This isn't because there is a specific intended referent; we don't really need one, as long as we are speaking a language in which all the systems of natural numbers are indiscernible. Whether this view 'eliminates the natural numbers' depends what one means by 'eliminate'. The view maintains that 'the natural numbers' has a complete semantic role, but it denies that 'the natural numbers' is a singular term in *exactly* the sense of 'JC'.

2. Types and Terms

Now I sketch the formal development of MLTT (following Chapter 1 of *UFP* and Martin-Löf (1984)). Most of the ideas are already on the table. However, we need some grasp of the formalism to properly appreciate what is going on in HOTT, and in particular to understand the logical calculus. That will the subject of section 3.

Universes. I will only really need to refer to the first two universes, which I will denote by $\mathcal{U}$ and $\mathcal{U}'$. Each of these are types of types, in the sense that if A is a term of type $\mathcal{U}$, then we may also have terms of type A. The 'small' universe $\mathcal{U}$ is supposed to be an element of the 'large' $\mathcal{U}'$, and so, in the formal mode, $\ulcorner\mathcal{U}\urcorner$ is a term of type $\mathcal{U}'$. But $\mathcal{U}'$ is also supposed to *subsume* $\mathcal{U}$; this means that any term of type $\mathcal{U}$ also counts as a term of type $\mathcal{U}'$.

Types. As for the basic existence axioms, we are given constants

$$\mathbf{0} \qquad \mathbf{1} \qquad \mathbb{N}$$

of type $\mathcal{U}$, to denote the empty, unit, and natural numbers objects. And whenever A, B are terms of type $\mathcal{U}$ (or, mutatis mutandis, $\mathcal{U}'$), we have terms

$$A \times B \qquad A \sqcup B \qquad A \to B$$

again of type $\mathcal{U}$. (Notation: $A \to B$ corresponds to the mapping set B^A from categorical set theory. And in *UFP* coproducts are written as sums.) A small generalisation is in order: we will be allowed indexed products and coproducts. This means that when we have a type $F(a) : \mathcal{U}$ depending on $a : A$, we are allowed to form the product or coproduct of the $F(a)$ as a varies. We could have allowed such constructions in categorical set theory, but they would necessarily have been schemata.[5] They are more naturally expressed in a system with universes. To say that $F(a)$ depends on a is just to say that F is a term of type $A \to \mathscr{U}$; then

$$\prod_{(a:A)} F(a) \qquad \coprod_{(a:A)} F(a)$$

are also terms of type $\mathscr{U}$.

The method by which these types are characterised, however, differs significantly from the method used in categorical set theory. In CST we had a background theory — category theory — governing the primitive object and morphism sorts and the functions id, $\circ$. The specifically set-theoretic objects were characterised in terms of this primitive ontology by means of axioms expressed in first-order logic. Even the elements of objects were described in terms of this background ontology. In contrast, MLTT forgoes the background theory and even the background logic. Each term X of type $\mathscr{U}$ admits terms of type X, not to be analysed in terms of morphisms from a terminal object; the theory then characterises the basic objects by syntactic rules for term-formation. It will be easiest to see what this means by example.

The key example is provided by the function types. MLTT provides as a rule of syntax that given a term $A : \mathscr{U}$, one can form a term id_A of type $\ulcorner A \to A \urcorner$ (the identity function). Similarly, there is a rule that given a term a of type A and a term f of type $A \to B$, one can form a term $f(a)$ of type B. Contrast to the situation in CST: there, category theory provided, for each $A : \mathscr{O}$, a term id_A of type $\mathscr{M}_A^A$; the axioms governing mapping sets then entail that there is an appropriately corresponding element of any mapping set A^A. The fully-spelled-out explanation of function evaluation

[5] The existence of indexed coproducts is a version of the replacement axiom, which is a schema in first-order Zermelo theory.

would be even more complicated. In MLTT there is no place for morphism-types distinct from the mapping sets, and this is a great simplification.

It would be interesting to analyse how this method of characterisation works. However, that analysis would take me too far away from the main line of inquiry here. (In particular, a presentation of the full rules in logical order would require a discussion of identity, which I am saving for later.) What I will do is list enough of the term-forming rules to give a flavour of the theory and to serve my current narrow purposes. I hope, though, that it will be clear that these characterisations fundamentally agree with those given in CST. In particular, one should not get the impression that these characterisations are extensional. For example, we are told that any *term* of type $A \to B$ can be combined with a term of type A to form a term of type B. It does not follow that any *element* of $A \to B$, perhaps not given to us as such, is some kind of gadget for converting elements of A into elements of B. Indeed, in Chapter IV.6, I will deny that it is so.

2.1. Constructing Terms. To the rules, then. As we shall see, it is highly suggestive to write them in the style of rules of inference in a natural deduction systems. The 'premisses' or inputs are various known terms, and the 'conclusion' or output is a new one. For example, the already-mentioned rule of function evaluation can be represented thus:

(4) $$\frac{a : A \quad f : A \to B}{f(a) : B}$$

There are other rules, too, including one for forming terms of $A \to B$ by λ-abstraction; but I pass them over.

As for the singleton set **1**, we are given a name for its element:

(5) $$* : \mathbf{1}\,.$$

Of course, this element should be unique. But I do not want to make the detour required to say so formally. We expect that the empty set $\mathtt{0}$ has no elements, so we can hardly have a rule for forming terms of type $\mathtt{0}$. But it should be an initial object in the sense that it admits a unique function to every other set. For each $A : \mathscr{U}$ we are given a name for that function:

$$\text{init}_A : \mathtt{0} \to A. \tag{6}$$

For the product type $A_1 \times A_2$, we can introduce terms as ordered pairs:

$$\frac{a_1 : A_1 \quad a_2 : A_2}{(a_1, a_2) : A_1 \times A_2} \tag{7}$$

Conversely, from any term we can form terms for its components:

$$\frac{c : A_1 \times A_2}{\text{pr}_1(c) : A_1} \qquad \frac{c : A_1 \times A_2}{\text{pr}_2(c) : A_2}. \tag{8}$$

The indexed products generalise these binary products. Heuristically, in the case of binary products, we have a type A_i for each i in the index set $X :\equiv \{1, 2\}$; an element (a_1, a_2) of the product can be understood as a function on $\{1, 2\}$ whose value on i is a_i. (This is only a heuristic, because we do not at this point have such a set $\{1, 2\}$ at our disposal!) The indexed products generalise this to the case of an arbitrary index type X. In fact, the indexed products are usually called 'dependent function types', because their elements are given by functions on the index type, with the type of the value depending on the argument. They thus generalise both function types and product types.

The main rule for indexed products generalises both (4) and (7):

$$\frac{a : A \quad f : \prod_{x:A} F(x)}{f(a) : F(a)} \tag{9}$$

Explicitly, this generalises (4) because it says we can 'evaluate the function f at argument a'. But it also generalises (7) because it says we can 'project the tuple f to its ath component'.

The coproduct $A_1 \sqcup A_2$ has elements $\mathrm{in}_1(a_1), \mathrm{in}_2(a_2) : A_1 \sqcup A_2$ for each $a_1 : A_1$ or $a_2 : A_2$:

(10)
$$\frac{a_1 : A_1}{\mathrm{in}_1(a_1) : A_1 \sqcup A_2} \qquad \frac{a_2 : A_2}{\mathrm{in}_2(a_1) : A_1 \sqcup A_2}$$

This generalises to indexed coproducts. A term $\mathrm{in}_i(a_i) : A_1 \sqcup A_2$ could be written as an ordered pair (i, a_i). For a general index type X, we start from a function $A : X \to \mathscr{U}$. Terms of type $\coprod_{(x:X)} A(x)$ can be constructed as pairs (x, a_x) with $x : X$ and $a_x : A(x)$. This indexed coproduct thereby generalises both binary coproducts and binary products; sometimes it is called a type of 'dependent pairs'. Generalising (10) and (7), we have a rule:

(11)
$$\frac{x : X \quad a_x : A(x)}{(a, a_x) : \coprod_{a:A} F(a)}$$

We can also project a pair to its components, generalising (8):

(12)
$$\frac{s : \coprod_{x:A} F(x)}{\mathrm{pr}_1(s) : A} \qquad \frac{s : \coprod_{x:A} F(x)}{\mathrm{pr}_2(s) : F(\mathrm{pr}_1(s)).}$$

I will leave out any discussion of the natural numbers type; suffice to say that we are given terms $0 : \mathbb{N}$ and $\mathrm{succ} : \mathbb{N} \to \mathbb{N}$.

3. Logic

So far I have described what types HoTT postulates, and (partially) characterised them via rules for constructing and manipulating their elements. To give a fully formalised theory, we have to combine these postulates and rules with a system of predicate logic. (Why not just state the theory in natural language? One reason is that we expect to require strict type distinctions, which are difficult to enforce in natural languages.) One way to do this would be to incorporate the basic ontology into a background system of (say) first order logic. This would bring us back to categorical set theory à la FOLDS.

The strategy of dependent type theory, however, is slightly different. It postulates a type Prop of 'propositions' — roughly, truth values.[6] The move (which I will not seriously defend) goes back to Frege: we construe a closed sentence as a name for an element of Prop. Similarly, a predicate Ψ on A is a 'propositional function' $A \to \mathrm{Prop}$. The universal quantification of such a predicate should again name an element of Prop, so the quantifier $\forall(a : A)$ should be a higher order function $(A \to \mathrm{Prop}) \to \mathrm{Prop}$. Thus the strategy is to subsume the linguistic rules of sentence-formation under the general rubric of constructing elements of types. Since we consider $\mathscr{U}$ itself as a type in a larger universe $\mathscr{U}'$, we even have the freedom to quantify over propositional functions $\mathscr{U} \to \mathrm{Prop}$. The logic is higher-order rather than first-order in this sense.

What is Prop? In principle, any two-element set would do. According to HoTT, though, the elements of Prop are the sets with at most one element.[7] Thus Prop will be a subtype of $\mathscr{U}$. This answer looks strange, and

[6]In CST, Prop corresponds to the 'subobject classifier'. In *UFP* what I am calling 'propositions' are called 'mere propositions'.
[7]The discussion here will have to be informal; at the moment, we can't say formally what a 'at most one' means! I will clear this up a bit in ch. IV.2.

I can only partly motivate it. First, a standard heuristic is that the truth values $\bot$ and $\top$ are 'nullary relations' and so, in extension, subsets of 'the nullary product' $\mathbf{1}$. We could identify these subsets with $\mathbf{0}$ (as $\bot$) and $\mathbf{1}$ (as $\top$). But, second, it will turn out in HoTT that all empty sets (being mutually isomorphic) are equal to $\mathbf{0}$ and all singleton sets (being mutually isomorphic) are equal to $\mathbf{1}$. So we may as well say that Prop contains all sets with at most one element. A posteriori, this choice of Prop makes good on the following central idea: the sentence-forming rules *just are* the standard type-construction rules and the rules of inference *just are* the standard term-formation rules, as applied to propositions.

To see how this works, suppose that P, Q are propositions (again: sets with at most one element); then the set $P \to Q$ of functions is again a proposition. A moment's thought shows that $P \to Q$ has one element if and only if P is empty or Q is non-empty. It is naturally interpreted as the proposition that P implies Q. The function type $P \to \mathbf{0}$ is, similarly, the proposition that not P, and the product type $P \times Q$ is the proposition that P and Q. Moreover, universal quantification is a kind of infinitary conjunction, and indeed corresponds to an indexed product of propositions. That is, if $F : X \to \text{Prop}$ is a predicate, then its universal quantification should just be $\prod_{x:A} F(x)$. I will therefore adopt the abbreviations

$$\top :\equiv \mathbf{1}$$

$$\bot :\equiv \mathsf{0}$$

$$\neg P :\equiv P \to \mathsf{0}$$

$$P \wedge Q :\equiv P \times Q$$

$$P \vee Q :\equiv P \sqcup Q$$

$$\forall(x : X)P(x) :\equiv \prod_{x:X} P(x)$$

whenever it is convenient to stress these logical readings of the type-constructors. (If you are wondering about disjunction and existential quanification, I will get there in a minute!)

This explains what I meant by the claim that the sentence-formation rules just are the type-construction rules. For example, the rule that one can form the conjunction of two sentences just is the rule that one can construct the product of two types. I also claimed that the inference rules just are the term-construction rules. To see how this works, consider first the main rule for function types:

(4 again) $$\frac{a : A \quad f : A \to B}{f(a) : B}$$

A proposition holds just in case it is not empty. So, if A and B are propositions, this rule can be read as modus ponens: from the truth of A and the truth of $A \to B$ one can deduce the truth of B. In this connection, (5) asserts the True while the rule (6) is *ex falso quodlibet*. The rule (9) is universal elimination. And of course there are other rules.

In the discussion so far, I have neglected disjunction and existential quantification. They are definable in the usual way using negation, conjunction, and universal quantification. But one *might* expect an alternative. If we read coproducts as disjunctions, rule (10) looks like $\vee$-introduction; rule (11) looks like $\exists$-introduction. However, conceptually, something is amiss. On the one hand, we expect that $\top \vee \top = \top$; but we also expect $\mathbf{1} \sqcup \mathbf{1}$ to have two elements, not one. $\mathbf{1} \sqcup \mathbf{1}$ is not a proposition. This difficulty infects the term-formation rules, as seen most clearly in the case of rule (12). Interpreted as an inference, it would pass from $\exists(x : A).F(x)$ to some particular element $\mathrm{pr}_1(s)$ of A such that $F(\mathrm{pr}_1(s))$ is true. Of course, this is impossible for classical existential quantification; it is, though, what one would expect for *constructive* existential quantification, and I will come to that shortly.

If we are sticking to classical logic, then we cannot say that disjunction and existential quantification are special cases of coproducts. Still, we can see how they should be *related* to coproducts. Suppose that, for any type A, we could form a type $||A||$ — the 'propositional truncation' of A — which is a proposition, and which is empty if and only if A is. In other words, $||A||$ is the proposition that A is non-empty. Then, simply by counting elements, we expect to have

$$(13) \qquad P \vee Q :\equiv ||P \sqcup Q|| \qquad \forall(x : X).F(x) :\equiv \left\| \coprod_{x:X} F(x) \right\|$$

whenever P, Q, and all values of F are propositions. We might *also* guess that we can define $||A||$ to be the type $((A \to \mathbf{0}) \to \mathbf{0})$; if everything is working properly, this latter will have at most one element, and will be empty if and only if A is.[8]

[8]If $\ulcorner\neg A\urcorner$ is defined in general as $A \to \mathbf{0}$, we could write $\ulcorner\neg\neg A\urcorner$ for $||A||$. However, one may be wary of writing $\ulcorner\neg A\urcorner$ when A is not a proposition.

Let me summarise this discussion. I have relied on a certain amount of prior intuition to get to this point, but the end-result faithfully represents an orthodox view of HoTT. Among all types, there are some that are 'propositions' (and which are ultimately the sets with at most one element). The usual type-construction rules amount to a predicate calculus for these propositions, and the usual element-construction rules give inference rules. The caveat is that disjunctions correspond to coproducts only *up to* propositional truncation (13).[9]

Finally, a comment about comprehension. In my discussion of categorical set theory, I postulated a comprehension scheme, allowing one to form the 'extension' of a predicate on a set A. In HoTT, we get such a thing automatically. A 'predicate' on A is a term F of the function type $A \to \mathrm{Prop}$. The extension is then the dependent pair type $\coprod_{a:A} F(a)$, which we could even call $\{a : A \mid F(a)\}$. We would ordinarily introduce terms of this type as pairs (a, p) with $a : A$ and $p : F(a)$. But since each $F(a)$ has at most one element, such terms correspond to the $a : A$ *such that* $F(a)$ is true.

4. Against Intuitionistic Logic

It emerged in the last section that the construction rules of HoTT take the form of intuitionistic rules of inference. This is a remarkable fact. What should we make of it?

There is a radical line of thought, going back to Martin-Löf's original work, and the source of much equivocation in *UFP*. It holds that arbitrary types (and not just some special ones) count as propositions, that the

[9]I alluded to another caveat in the introduction. Namely, one still has to impose the law of excluded middle (LEM) as an axiom in order to get classical logic. LEM alternatively follows from a version of the axiom of choice (*UFP*, ch. 3). In CST the axiom of choice (part of my Function Plenitude axiom) plays a similar role: it guarantees that the 'internal logic' of the topos is classical. MLTT is essentially the internal logic of the topos of sets (cf. Awodey (2009)).

logical functors just are the type-constructors (without any sort of propositional truncation), and that the inference rules just are the intuitionistic ones. According to this line of thought, an instance (4) of 'modus ponens' with the 'conclusion' $f(b) : B$ asserts (for some sense of 'asserts') that $f(b)$ is a proof of the proposition B. The intuitionistic character of the logic is explained in terms of what counts as a proof of a proposition of a certain form (Martin-Löf, 1984, p. 12). For example (the story goes), a proof of $A \sqcup B$ is either a proof of A or a proof of B; so, from (the proof of) the former one can deduce (a proof of) a determinate one of the latter.

However, I claim that this radical line of thought is, at best, conceptually superfluous. While it affirms that types are propositions, it is *also* committed to the view that types are collections: the purported quantifiers $\prod$ and $\coprod$ bind variables that range over elements of arbitrary types. My antecedent concern in this essay is also with types as 'collections' rather than 'propositions'. But, once we admit 'types as collections' then the talk of 'types as propositions' adds nothing but mystification. To give an historical analogy, Frege (1893, §2) recognised propositions (truth values) as objects on a par with all others. He even defined the logical functors to apply to any objects, in line with his principle that a function must accept any object as an argument. In that sense, objects generalised propositions. Crucially, though, his judgment stroke was always affixed to the horizontal stroke that converted arbitrary objects into truth values. Because of this, it would be wrong to ascribe to Fregean objects in general the logical status of propositions.

The same goes here: if A is a collection, what does it mean to assert that A? What proposition is the natural numbers? What is 'a proof of the natural numbers'? The 'types as propositions' view in fact forces the answer: for any type A, there is a valid biconditional

$$A \leftrightarrow \coprod_{a:A} \top.$$

(That is, one can construct functions in both directions.) According to the reading of coproducts as existential quantifiers, the right-hand side of the biconditional represents the proposition that A is non-empty. So too, A must represent the proposition that A is non-empty. Evidently, then, an element of A is 'a proof of A' just in the sense that it indicates that A is non-empty. But we need not abandon classical logic to recognise that one can conclude, given an element of A, that A is non-empty. This inference is given by $\exists$-introduction:[10]

$$\frac{a : A}{||a|| : \exists(a : A).\top}$$

There is no conceptual reason to call A 'a proposition' and a 'a proof of A'. To be sure, it is a beautiful observation that, because each element of A indicates that A is non-empty, certain standard methods for constructing elements of types run in parallel to certain standard inference rules. But why insist on saying more?

Let me make clear why this critique does not impugn my preferred view that types with at most one element represent propositions. Only for such a type A does constructing an element coincide with proving that A is non-empty; for if A is known to be non-empty one can construct its element by definite description. Moreover, by definition, such a type A is

[10]The displayed inference is, more pedantically, from $a : A$ and $* : \top$. Note the conclusion that A is non-empty can be represented by the displayed existential quantification or by the propositional truncation $||A||$; they are logically equivalent and (in HoTT) equal.

non-empty if and only it represents the True. For these types, and only for these types, is construction naturally interpreted as proof.

And yet, having said all this, once we are clear on what we are doing, it is perfectly harmless to *talk like a constructivist*. I will occasionally do so, in a well-marked way. For example, if I call P a 'constructive proposition' I just mean that P is a type. Do not imagine that a constructive proposition is genuinely a proposition. If I say that P is 'constructively true', I just mean that there is some particular element of P in the offing. Do not imagine that the non-proposition P is genuinely true. If I say that P 'constructively implies' Q, I just mean that we have some particular function $P \to Q$; do not imagine that constructive implication is a form of genuine implication. Once one starts, it is easy and useful to go on like this. I will occasionally indulge, but my general scepticism should be clear.

5. Conclusion

Martin-Löf's dependent type theory has some advantages over categorical set theory. Its universes and constants better reflect mathematical practice. It might appear that the use of constants — for example, a specific term $\mathbb{N}$ for the natural numbers — forces us into an (if anything) object-structuralist interpretation of MLTT. I argued in section 1.2 that the situation is not clear-cut. I sketched an idea of how singular terms might work in categorical set theory without forcing us into metaphysical commitments.

The other main novelty of MLTT is the use of a type of propositions. There is a conservative version of this move, which preserves classical logic; but there is also a radical intuitionistic one, according to which every type represents a proposition. I argued that this radical version is conceptually spurious, even as it may provide a convenient way of speaking.

CHAPTER IV

Homotopy Type Theory

In Chapter II, I argued that the structuralist aspirations of categorical set theory are best fulfilled by a language that eschews the identity predicate for sets. From this point of view, the fundamental novelty of HoTT is that it reinstates this predicate. The way it does so is what led Awodey (2014) to declare that in homotopy type theory 'mathematical objects simply are structures'. Good news, then, for the object structuralist; but there is bad news, too, and the identity predicate of HoTT raises subtle interpretive issues. Those issues are the subject of this chapter.

Here is the plan. In section 1 ('The Good News') and section 2 ('The Bad'), I lay out the most conspicuous features of equality in homotopy type theory. The bad news is that the spurious intuitionism of ch. III.4 once more rears its ugly head. This leads to a face-value reading of HoTT according to which the logic of identity is intuitionistic. In section 3 ('The Constructive Identity View') I try to give this reading a fair hearing, but it must ultimately be rejected. In section 4 ('The Mere Identity View') I show how to do so. In section 5 ('Object Structuralism…') I revisit the problems raised in Chapter II; how exactly is it that HoTT manages to have object identity predicates? I argue that the key is *intensional typing*. But intensional typing is a bit mysterious; it goes with a weakening of the indiscernibility of identicals. In section 6 ('Intensional Typing and Indeterminate Reference') I sketch a way in which an object-structuralist might try to explain intensional typing. Finally, in section 7 ('Structuralism Without Equality?') I argue that there is unlikely to be a good account of identity, distinct from the 'Mere Identity View', within the framework of homotopy type theory.

1. The Good News

In Chapter II, I set aside the idea of imposing a skeletal axiom of the form

$$A \cong B \to A = B. \tag{14}$$

The reason for setting this aside was that it could not, by itself, solve the Object Identity Problem. However, the skeletal axiom may have independent motivation within a structuralist programme. Roughly, the skeletal axiom corresponds to the motto that 'isomorphic systems have the same structure'. However, a bit of care is needed here. Notice that there is no overt system/structure distinction in the skeletal axiom; it says that isomorphic *objects* are the same *objects*. But suppose we accept the object-structuralist thesis that each structure has 'places' that themselves exemplify the structure. The skeletal axiom can be taken to express the idea that structures are equal if they have isomorphic systems of places. This, I take it, is the structuralist motivation for the skeletal axiom.

Homotopy type theory imposes an axiom of the form (14). It is called UNIVALENCE. It is asserted of sets, but it turns out — I think it fair to say *miraculously* — to apply to more complicated structures as well. Isomorphic sets are equal, but so are isomorphic groups, or isomorphic fields. (However, the phenomenon is not completely general! I will mention some counterexamples in Chapter V.) For future reference, let me explain more carefully what this means. (The impatient reader may skip to the next section and refer back as necessary.)

Structure-Types. How can we talk about the type of all groups, say, as a mathematical object? (Remember Hellman's worry in ch. II.1.1.) The existence postulates of MLTT allow us to do so. Here is a fairly general version of the story. Suppose we are interested in models of a theory stated in a standard (first- or second-order) language with relation symbols

$$R_1, \ldots, R_m$$

of arity $a_1, \ldots, a_m$, and function symbols

$$f_1, \ldots, f_n$$

of arity $b_1, \ldots, b_n$. Let Ψ be the conjunction of the finitely many axioms. A model of this theory is an $(m+n+1)$-tuple

$$(X, R_1, \ldots, f_1 \ldots)$$

consisting of a set X and some relations and function on it, satisfying Ψ. Such tuples are naturally understood as elements of the type[1]

$$\coprod_{(X:\mathcal{U})} \coprod_{(R_1:X^{a_1}\to \mathrm{Prop})} \cdots \coprod_{(R_m:X^{a_m}\to \mathrm{Prop})} \coprod_{(f_1:X^{b_1}\to X)} \cdots \coprod_{(f_n:X^{b_n}\to X)} \Psi(X, R_1, \ldots, f_1, \ldots).$$

Here X^a is the product of X with itself a times. I will say that a type presented in this form is a *structure-type*.

[1]Strictly speaking, a term of the displayed type is an $(m+n+2)$-tuple $(X, R_1, \ldots, f_1, \ldots, *)$, where $*$ is an element of the proposition $\Psi(\cdots)$. But the datum $*$ can be interpreted as the datum that $\Psi(\cdots)$ is true.

To give the simplest examples, $\mathscr{U}$ itself (or equivalently $\coprod_{X:\mathscr{U}} \top$) is the structure-type of sets.[2] Similarly, a *pointed set* is a pair (X, x) with $x : X$, and

$$\text{PSet} :\equiv \coprod_{(X:\mathscr{U})} \coprod_{(x:X)} \top$$

is the structure-type of pointed sets. We similarly have the structure-type of groups, the structure-type of natural-numbers systems, and so on. It is also easy to generalise this basic notion of 'structure type' to include structure-types of models of (say) many-sorted or dependently-sorted theories. So we have a structure-type of categories, too.

If we have a structure-type $\mathscr{T}$, with terms A, B, it is easy enough now to define a type $\text{Iso}_{\mathscr{T}}(A, B)$ of isomorphisms between A and B in the expected sense. For example,

$$\text{Iso}_{\text{PSet}}(X, x; Y, y) = \coprod_{f:X \to Y} [f \text{ has an inverse }] \wedge f(x) = y.$$

What Univalence implies (*UFP*, ch. 9.8) is that there is a function

$$\text{Iso}_{\mathscr{T}}(A, B) \to A = B.$$

Given my motivation for the skeletal axiom, it might be clearer to think of the elements of the structure-type of groups (say) as *group-structures* rather than mere groups; but I will normally just call them groups, like everybody else.

[2]As will appear in the next section, talk of 'sets' in this whole discussion should be taken informally; some elements of $\mathscr{U}$ will not be sets in a technical sense. That leads to some further, mainly irrelevant subtleties; see for example fn. 6.

2. The Bad News

Out with it, then: the 'identity predicate' of homotopy type theory is not a predicate at all.

In light of the discussion of intuitionistic logic in ch. III.4, the reader can anticipate what I mean. The identity of A with B is (apparently) represented by a type $A \overset{**}{=} B$, but this type may not be a proposition in the sense of ch. III.3. (I use the symbol $\overset{**}{=}$ to distinguish this 'predicate' from the version I will favour.)

What should we make of this? That is the central question that any interpretation of HoTT must face. Until now I have been coy about what exactly a proposition is in homotopy type theory; I have said that it is 'a set with at most one element', but that characterisation relies on a theory of identity. To get a grip on the issues at hand, we ought to look a *little* more closely at how the theory of identity and the notion of proposition are supposed to hang together. Then I will turn to questions of interpretation.

Identity, Propositions, and Sets. For any $A : \mathcal{U}$ there will be a function $\overset{**}{=} : A \times A \to \mathcal{U}$; thus, for any $a, b : A$, $a \overset{**}{=} b$ is a type. (The same goes for the larger universe $\mathcal{U}'$ in place of $\mathcal{U}$; so we could be talking about the identity 'predicate' on $\mathcal{U} : \mathcal{U}'$.) From what I have said, 'X is a proposition' should correspond to the type[3]

$$\mathrm{isProp}(X) :\equiv \prod_{(x,x'):X\times X} (x \overset{**}{=} x').$$

[3]Recall that $\prod$ corresponds to universal quantification. I resist writing '$\forall$' here because the formula being bound is not necessarily a proposition.

It will come out in the wash that isProp(X) is itself a proposition (*UFP*, ch. 3.3). The type Prop then consists of the types that are propositions:

$$\mathrm{Prop} :\equiv \coprod_{(X:\mathscr{U})} \mathrm{isProp}(X).$$

Thus, strictly speaking, an element of Prop is given by a pair (X, p) with $X : \mathscr{U}$ and $p : \mathrm{isProp}(X)$. But (again) I interpret the datum p as the datum *that* X is a proposition.

Coming full circle, we could take it as an axiom that equations are always propositions:[4]

(15) $$\forall(A : \mathscr{U}).\forall(a, a' : A).\,\mathrm{isProp}(a \overset{**}{=} a').$$

This is what homotopy type theory neglects to do. In fact, the theory defines 'a set' to be a type A of which

(16) $$\forall(a, a' : A).\,\mathrm{isProp}(a \overset{**}{=} a')$$

is true; so the denial of (15) is exactly the mechanism by which HoTT claims to generalise set theory.

This is where the skeletal Univalence axiom enters in. For A and B in $\mathscr{U}$, the axiom is actually given by an *isomorphism*

(17) $$\mathrm{Iso}(A, B) \to A \overset{**}{=} B$$

[4]Taking a proposition as an axiom means that one assumes given a term of the corresponding type.

between the type of isomorphisms and the identity type.[5,6] Isomorphism is isomorphic to identity. Since in general there will be *lots* of isomorphisms between A and B, this means that $A \overset{**}{=} B$ may not be a proposition. (And therefore at least the universe $\mathscr{U}$ is not a set.)

3. The Constructive Identity View

One possibility is that we should take the situation at face-value; HoTT demands a constructive reading of identity claims. It is a hard to state precisely what this means (unless we are to let the formalism speak for itself). But since it is the face-value interpretation, it deserves a fair hearing. Let me try.

As a first step, let me explain the analogue for isomorphism rather than identity. Suppose that A and B are isomorphic sets. The mere statement that A and B are isomorphic, as usually understood, does not allow us to pass from elements of A to elements of B. However, there is a different 'constructive' understanding according to which 'A and B are isomorphic' at least implies that one has in mind a particular isomorphism between A and B. On this constructive reading of isomorphism claims, one can pass from elements of A to elements of B using the implied isomorphism. The logic of isomorphism in this constructive sense is bound to be intuitionistic. For example, 'A and B are not isomorphic' presumably still means that there is no isomorphism at all between them. But it is then not true that either A and B are not isomorphic or A and B are isomorphic [and one has a particular isomorphism between them]. The law of excluded

[5]As one might expect, the inverse 'implication' $A \overset{**}{=} B \rightarrow \mathrm{Iso}(A,B)$ can be constructed directly from a version of the indiscernbility of identicals (see §5).

[6] Note that the definition of 'isomorphism' (or 'equivalence', as it is usually called in HoTT) itself involves equality: a function $f : A \rightarrow B$ is an isomorphism if there exists a function $g : B \rightarrow A$ such that $f \circ g = \mathrm{id}_B$ and $g \circ f = \mathrm{id}_A$. This leads to some technical subtleties; but the basic point is to ensure, by propositional truncation, that being an isomorphism is a genuine predicate.

middle thus fails. On the face of it, homotopy type theory takes a similar constructive attitude towards identity claims.

But there is an obvious disanalogy between identity and isomorphism. The story about isomorphism rested on a distinction that does not seem to have an analogue for identity — namely, the distinction between the mere fact of isomorphism and the existence of a *particular* isomorphism. What in the world is 'a particular identity'? The picture may become *slightly* less baffling if we remember that, according to the skeletal axiom, isomorphism and identity are equivalent. Remember that Univalence is given by an isomorphism (17) between the type of isomorphisms and the identity type. So one formulation of the view is that isomorphism, constructively understood, is the right notion of 'identity between structures'. On this formulation, to 'assert' that structures are equal just is to give an isomorphism between them.

So much for my attempt to *state* the view. What about its motivation? Part of what is driving the view, I suspect, is a mixed bag of notational, historical, and sociological accident. Moreover, the aim of using HoTT to do homotopy theory is a confounding factor (for more on which, see §4 and ch. V.2). However, the view is undeniably *also* driven by a remarkable feature of contemporary mathematical practice, which is hard to entirely discount. It is probably best to proceed to an example of a kind ubiquitous in modern geometry. Suppose I have a manifold M, which is the union of two overlapping sets $M = U \cup V$. Suppose I have a (real-valued) function f on M. It determines, by restriction, functions f_U on U and f_V on V. Conversely, functions f_U and f_V determine a function f, *given only that they are equal* on $U \cap V$. Now modify the scenario. Suppose I have a vector bundle f on M. It determines, by restriction, vector bundles f_U on

U and f_V on V. Conversely, vector bundles f_U and f_V determine an f, *given a particular isomorphism* between them on $U \cap V$. So here is the point: a mere identity in the first scenario has as its analogue in the second scenario the datum of a particular isomorphism. There is clearly a lot one could say here, but this sort of example does provide prima facie evidence from mathematical practice for the view that, sometimes, constructive isomorphism claims are the appropriate generalisation of plain-old identity.

As I said, the view is difficult to state precisely; relatedly, it would be hard to argue for it briefly. But let me tell a story about how one might come to accept it naturalistically. First, experience might lead one to accept that contemporary mathematics rarely talks about identity between systems — but it does construct and keep track of particular isomorphisms between them. Ubiquitous examples like that above then show how this mathematical practice is well represented as a kind of constructive reasoning about isomorphism which is parallel to classical reasoning about identity, and indeed can be seen to generalise it. Then one recognises that this informal reasoning can be perspicuously formalised in a system with a unified treatment of isomorphism and identity. And then one doesn't look back. It just comes out (the story goes) that the best formalisation of mathematical practice has an identity/isomorphism predicate with intuitionistic inference rules, and that this is the closest thing there is to the classical identity predicate. The philosopher ought to respect these facts on the ground.

4. The Mere Identity View

But this *must* all be misguided, if it asks us to construe $A \overset{**}{=} B$ as a proposition. I rehearsed the reasons in ch. III.4.

In the case of isomorphism, we recognize *both* the collection $\mathrm{Iso}(A,B)$ of isomorphisms *and* the proposition that A and B are isomorphic; there is no conceptual reason to conflate them. So too, we are bound *in any case* to recognize $A \stackrel{**}{=} B$ as a collection, and at that point it is just a matter of confusion to imagine that this collection is 'a proposition' or to imagine that either its elements or constructions of those elements are 'proofs'. At best, we can *talk* like a constructivist, in the way I describe on p. 71. Indeed, I will refer to $\stackrel{**}{=}$ as **constructive identity.** The only assertable proposition in the neighbourhood of $A \stackrel{**}{=} B$ is the *non-emptiness* of that type. If anything, then, it must be the propositional truncation

$$A = B :\equiv \left\| A \stackrel{**}{=} B \right\|$$

that represents the identity of A and B. I will call this predicate **mere identity**.[7] There is no way around this conceptual point. The example from mathematical practice only goes to show that it may be convenient to talk like a constructivist from time to time. To take that talk at face value is confused.

In any case, the evidence from mathematical practice is ambiguous. For example, people have studied the consistency of HoTT relative to Zermelo set theory, by interpreting HoTT in ZFC (Kapulkin et al., 2012, e.g.). But when they do so, they do not interpret $\stackrel{**}{=}$ as identity. (*Obviously* not. In set theory, identity is a genuine proposition.) These seemingly deviant interpretations are not only of interest in regards to formal consistency. As I explained in the introduction, one of the motivations for developing HoTT is the study of homotopy theory. Now, homotopy theory, as

[7]In *UFP* '$\stackrel{**}{=}$' is simply written '$=$' and called 'propositional equality'. This name is seriously misleading. In that book, mere identity does not have a name, but what I call 'propositions' are there called 'mere propositions'.

a standard field of mathematics, has been developed in set theory, and the interpretations I have mentioned essentially translate between homotopy theory à la HoTT and homotopy theory à la set theory.[8] So, to the extent that set-based homotopy theory is correct, it looks as if even the *intended* interpretation of constructive identity is not identity.

Of course, the reason for doing homotopy theory with HoTT is that its development in set theory is somehow infelicitous. Probably, then, we should not take the set-based interpretation of HoTT as literally the intended model. Still, the point remains that the interpretation of constructive identity in HoTT is, to a significant extent, up for grabs.

So, on my preferred view, the *proposition* $A = B$ represents the equality of A and B. The 'good news' of section 1 is not impugned; my discussion there can be applied to mere identity. But where does this leave the collection $A \overset{**}{=} B$? The answer has two parts. First, when A and B are elements of a *set* (in the specific sense of §2), mere equality and constructive equality coincide. That is just what it meant to be a set; in this most important case, there is nothing to explain. The second part of the answer is that between sets, we have Univalence, and we can eliminate constructive equality in favour of isomorphism. Univalence gives an isomorphism between $\mathrm{Iso}(A, B)$ and $A \overset{**}{=} B$. Constructively speaking, $\mathrm{Iso}(A, B)$ even *equals* $A \overset{**}{=} B$. We can use that to eliminate the mysterious elements of $A \overset{**}{=} B$ in favour of the much less mysterious elements of $\mathrm{Iso}(A, B)$. I will discuss this eliminitive strategy more in section 7. It works, as far as it goes. But it does not go all the way.

[8] Here I am referring to 'set theory' in general as opposed to specifically Zermelo set theory. The basic picture would remain even if we were to develop homotopy theory within the set-theoretic fragment of HoTT. The basic idea of this homotopy interpretation is that each type is interpreted as a topological space. A term $e : a \overset{**}{=} b$ is interpreted as a continuous path between the points a and b.

Why not? The problems are of roughly two kinds. First, there are problems that have a direct impact on the structuralist interpretation with which I have been concerned. The pre-eminent problem of this kind is related to the indiscernibility of identicals. That will be the main focus of the rest of this chapter.

Problems of the second kind bear on how we should interpret the way in which HoTT generalises set theory to homotopy theory. When we do so, we will be bound to consider constructive identities $A \overset{**}{=} B$ in which A and B are neither types nor elements of sets. Then we cannot explain $A \overset{**}{=} B$ by appealing either to Univalence or to the coincidence of constructive and mere identity. For the most part, I am just ignoring such problems for the purposes of this paper. But let me conclude this section with a few examples of what is at stake. Even if they do not bear on the structuralist interpretation per se, they may illuminate the picture as a whole.

Example 1. Sets and Propositions. In Chapter III.3 I said that a proposition was a set with at most one element. There was at least a heuristic basis for this claim. But then, when I got around to formalising things in §2, I defined the property of being a proposition using the *constructive* identity predicate:

$$\text{isProp}(X) :\equiv \prod_{(x,x'):X\times X} (x \overset{**}{=} x'). \tag{18}$$

Because of this, there will be types X that are *not* propositions, but of which

$$\forall(x, x' : X).(x = x') \tag{19}$$

is true. This seems to be a way in which $\overset{**}{=}$, and not $=$, plays the identity role. It is something to be explained. But my contention is that we need not construe $x \overset{**}{=} x'$ as a proposition in order to explain it.

I will say a little more along those lines after the next example; but note, in any case, that the issue does not arise *for sets*. For sets, constructive and mere identity coincide. Propositions are *sets* that satisfy (19). That is why the issue only bears on the generalisation of set theory.

Example 2: Function Extensionality. *Function extensionality* is the principle that if functions are point-wise equal (i.e they take the same values everywhere) then they are equal simpliciter. How should one formalise function extensionality in HoTT? If mere equality is equality simplicter, then function extensionality should be represented by the proposition

$$\forall(f, g : A \to B).\big(\forall(x : A).(f(x) = g(x))\big) \to (f = g). \tag{20}$$

However, this turns out to contradict Univalence. Rather, Univalence constructively implies the constructive analogue of (20):

$$\prod_{(f,g:A\to B)} \big(\prod_{(x:A)} f(x) \overset{**}{=} g(x)\big) \to (f \overset{**}{=} g). \tag{21}$$

The thrust of the example is that to preserve function extensionality, we must acknowledge that it is represented by the type (21) rather than (20). To that extent, at least, constructive equality plays the role of equality simpliciter.

Here is the rough idea of why (20) fails. Naively, if f is a function then we know

$$a = b \to f(a) = f(b). \tag{22}$$

But in HoTT the constructive version is also constructively true:

$$a \overset{**}{=} b \to f(a) \overset{**}{=} f(b). \tag{23}$$

That is, not only does f map merely equal arguments to merely equal values, it must also determine another whole family of functions that transform constructive equality into constructive equality.[9] The values of these additional functions can vary (with respect to mere equality) even though the values of f do not. And that is why f is not determined by its values in the sense of (20).[10]

Again, it is important to realise that the expected version (20) of function extensionality holds as long as A and B are sets. When A and B are sets, (21) and (20) coincide. So no one is denying the usual principle of function extensionality in the usual cases. Thus the distinction between (21) and (20), like that between (18) and (19), is a manifestation of the way in which homotopy type theory generalises set theory. All that is at stake here is how best to characterise this generalisation.

On the radical view of the previous section, the generalisation consists in recognising that the logic of identity between elements of non-sets is intuitionistic. In the current discussion, the evidence in favour of this view is that (23) looks just like the usual principle of functionhood and (21) looks just like the usual principle of function extensionality. We should accept that these formulae not only name types but also express these standard principles as propositions. Similar comments hold for Example 1.

[9] For types a and b, this means that f is a functor: it maps each isomorphism to another.
[10] Remarkably, functions can still be defined by λ-abstraction. That is roughly because λ-abstraction determines both the extension and the intension of the function.

On the other hand, I say that these types *can't* represent the usual principles, because they are collections and not propositions. Their orthographic resemblance to the usual principles doesn't change that. I say that the logic of equality is classical; the generalisation of set theory consists in recognising that there is more to individuation than equality. If 'individuation' sounds weasily, I would be happy to say that there is more to equality than *mere* equality. That is potentially a great discovery, and I do not wish to play it down. What I protest against is the thought that this something more is *propositional*. For example, perhaps we should gloss (23) by saying that f 'translates the whole structure of individuation.' We can thus try to escape constructivism, even as a way of speaking.

5. Object Structuralism and Intensional Typing

I have argued that we must consider mere identity, and not constructive identity, to be identity simpliciter. We can now look more closely at how mere identity in homotopy type theory fits into the structuralist perspective of this paper.

Here is the basic problem. I explained in §1 that, on the face of it, the Univalence axiom vindicates object structuralism. (What I said there goes through perfectly well with mere identity: isomorphic objects are merely equal.) Given the whole dialectic of this thesis, that vindication should appear deeply mysterious. The whole point of adopting a dependently typed language was to avoid equations between sets. And the whole point of avoiding equations between sets was to vindicate property structuralism. Doesn't *reintroducing* such equations just *reintroduce* the problems I considered in Chapter II? Don't we *know* that this can't work?

Intensional Typing. The basic answer is that it depends on how the object-identities interact with the typing system — roughly, on whether type declarations are extensional contexts. I will say that typing is *extensional* if the following condition holds: if we have a term $\ulcorner x \urcorner$ of type X, and a term $\ulcorner y \urcorner$ of type Y, and $X = Y$, then there is a fact of the matter whether x equals y. If typing is not extensional, I will say it is *intensional*.

Let me motivate the terminology. At a first pass, when I say 'there is a fact of the matter whether x equals y', I just mean that the equation $\ulcorner x = y \urcorner$ is well-typed. The equation is well-typed just in case x and y have the same type. But 'the same' in what sense? If it suffices that $\ulcorner X \urcorner$ and $\ulcorner Y \urcorner$ corefer, then typing is extensional. But if $\ulcorner X \urcorner$ and $\ulcorner Y \urcorner$ must have the same intension (or even the same syntax), then typing is intensional. Having said all that, a well-typed equation is only *sufficient* for there to be a fact of the matter whether x equals y. It might have been a necessary condition if the typing system ideally reflected the metaphysics. But, less ideally, the fact that $\ulcorner x = y \urcorner$ is badly typed might just mean that we are formally unable to express the intended fact, or we might have to express it by other means. With this problem in mind, here is a slightly broader sufficient condition for extensional typing. Typing is extensional if any term of type X is also, or can be converted into, a term of type Y, just on the basis that $X = Y$.

Here is why the distinction matters.[11] A *pointed set* is a set A along with a distinguished element a. The structure-type of pointed sets is thus $\mathrm{PSet} :\equiv \coprod_{(X:\mathcal{U})} \coprod_{x:X} \top$ (cf. §2). An isomorphism between pointed sets is a bijection that maps the distinguished element of the domain to the

[11] In Chapter II, I considered an example involving spans rather than pointed sets. However, having escaped from the rigid confines of CST, the following example is both simpler and more enlightening. Needless to say, the same kind of analysis applies to the earlier example.

distinguished element of the codomain. It is then easy to see that pointed sets (A,a) and (B,b) are isomorphic if and only if A and B are. Ideally, a skeletal axiom would lead us to say that the pointed sets (A,a) and (B,b) are equal. That is, in fact, the picture that emerges from univalence. But in this informal discussion, we *at least* want (A,a) and (B,b) to be *indiscernible*. That means: any predicate F on PSet must take the same value on (A,a) as on (B,b). As a specific example, we must have $F(X,x)=F(\mathbb{N},\mathtt{0})$ whenever X is countably infinite. But here is an apparent counterexample:

(24) $$F(X,x) \iff X=\mathbb{N} \text{ and } x=\mathtt{0}.$$

If this definition makes sense, then $F(\mathbb{N},\mathtt{0})$ is true but $F(\mathbb{N},1)$ is false. Intensional typing denies that the definition makes sense: it denies that there is a fact of the matter whether $x=\mathtt{0}$, even given that $X=\mathbb{N}$. This is a bit different from the strategy I used in ch. II.3. There, in a similar example, I noted that we can't *formally* define F by (24): the equation $\ulcorner x=\mathtt{0}\urcorner$ is badly typed, since x has variable type X and $\mathtt{0}$ has constant type $\mathbb{N}$. That sort of move sufficed when we were only interested in the properties definable in our first-order language. But once we have moved to the higher-order setting of MLTT, we can quantify over properties; the question is whether the admittedly informal definition of F picks out something in the domain of quantification.

To fend off an obvious objection: no one denies that there is a fact of the matter whether $\mathtt{0}=1$ as elements of $\mathbb{N}$. This might seem enough to distinguish $(X,x):\equiv(\mathbb{N},1)$ from $(Y,y):\equiv(\mathbb{N},\mathtt{0})$. But we have to distinguish two claims. First, the claim that typing is extensional: if $X=Y$, then there is a fact of the matter whether $x=y$. Second, the weaker claim that if X and Y are *given in the same way* then there is a fact of the matter whether

$x = y$. Intensional typing contradicts the first, but not the second of these claims. The view contends that the truth of $x = y$ is settled by fact that $\ulcorner X \urcorner$ and $\ulcorner Y \urcorner$ express the same intension, rather than by the fact that $\ulcorner X \urcorner$ and $\ulcorner Y \urcorner$ corefer. If that's right, we cannot parlay the distinction between 0 and 1 into a propositional function F on PSet. For example,

$$X \text{ is given to us as } \mathbb{N} \text{ and } x = 0 \tag{25}$$

is not a well-defined propositional function of (X, x).

What to make of all this? I characterised intensional typing as denying that there is a fact of the matter whether $x = y$, as if their identity were somehow indeterminate. No doubt I could have put it differently, but I think this formulation points the right way. Gareth Evans (Evans, 1978; Lewis, 1988) famously argued that apparently indeterminate identity must really be due to indeterminate reference. That is exactly the kind of interpretation I will suggest here. More precisely, I suggest how one might relate intensional typing to the problem of reference in object structuralism.

5.1. The Indiscernibility of Identicals. Before proceeding to those questions of interpretation, it may be interesting to see how intensional typing manifests itself in the formal system of HoTT. The main rule governing the identity predicate $X = Y$ is the indiscerniblity of identicals. Suppose that P is a predicate on $\mathcal{U}$, i.e. a term of type $P : \mathcal{U} \to \text{Prop}$. Then the rule is

$$\frac{e : X = Y}{e_* : P(X) \to P(Y).} \tag{26}$$

That is, if $X = Y$ then $P(X)$ implies $P(Y)$. Now, naively, this term-forming rule should hold for an *arbitrary* function $P : \mathcal{U} \to \mathcal{U}$, not just when the values of P are propositions. To take the most important case, suppose $P(X)$ is just X and $P(Y)$ is just Y. Then the generalised rule would allow us to construct the identity function $e_* : X \to Y$. *That would contradict intensional typing*: the equation $e_*(x) = y$ would express identity between $x : X$ and $y : Y$.[12] More generally, if $X = Y$ we must have $P(X) = P(Y)$; intensional typing prevents us from constructing 'the identity function' $P(X) \to P(Y)$. Thus, at a formal level, the basic symptom of intensional typing is that the identity of indiscernibles (26) holds *only* for propositional functions.

In contrast, the constructive version of (26) is constructively valid. For any $P : \mathcal{U} \to \mathcal{U}$, we have a rule[13]

(27) $$\frac{e : X \overset{**}{=} Y}{e_* : P(X) \to P(Y).}$$

One might say that the cost of giving up the radical view of constructive identity is that we must give up the full strength of the indiscernibility of identicals. But one should feel no regret. The rule (27) isn't *really* the indiscernibility of identicals, nor can it restore extensional typing. The type $X \overset{**}{=} Y$ is relevant, and important, and so on, but it doesn't express a proposition. For example, the naive thought is that e_* is the identity map between $P(X)$ and $P(Y)$. But it can't be that straightforward: e_* depends on e. If we had some other term $f : X \overset{**}{=} Y$ in the offing, it would be ambiguous which of e_* and f_* was supposedly the identity map. It is better simply not to play the game.

---

[12]In particular, we could formalise (24) by
$$F(X, x) :\equiv \exists(e : X = \mathbb{N}).e_*(x) = 0.$$

[13]This is a special case of a rule called PATH INDUCTION, which is ultimately doing a lot of the work; but I will not need to invoke it explicitly.

6. Intensional Typing and Indeterminate Reference

Structural objects, or ante rem structures, are supposed to have elements or 'places' that themselves exemplify the structure in question. So, for example, the natural numbers structure has countably many places that are ordered in a certain way. The individual natural numbers are supposed to be these places; the number five, for example, is one of them. On the other hand, a structure is supposed to be what isomorphic systems have in common. In the style of Fregean abstraction:

$$\text{Structure}(X) = \text{Structure}(Y) \iff X \cong Y.$$

Then the structure of X is something like the isomorphism class of X. But isomorphism classes do not have elements; at least, not in the same sense that systems do. As Burgess (1999, p. 286) put it in his review of Shapiro (1997), we can define a 'direction' to be an equivalence class of lines (where parallel lines are equivalent); but a direction does not contain points in the same way lines do. So it is a bit mysterious where these 'places' come from.

Nonetheless, we seem to be able to talk about places. Crucially, we tend to refer to them in a specific way. Given an element of a system, we can talk about the place occupied (or the role played) by that element in the structure. That seems to be the basic way we can think about and refer to places. Again, I am not particularly interested in *defending* object structuralism. But if we are willing to entertain this picture, it seems to shed some light on intensional typing.

How so? Well, suppose I have two models of the complex numbers; call them X and Y. Suppose that the two square-roots of -1 in X are I and $-I$, and those in Y are J and $-J$. Naively, I can refer to the place in the complex numbers structure occupied by I; call this $P(I)$. I can also refer to the place in the complex numbers structure occupied by J; call this $P(J)$. Does $P(I)$ equal $P(J)$? Of course, two complex numbers are equal or they aren't. And yet there seems to be no fact of the matter whether the place occupied by I is the same as the place occupied by J. For there are some isomorphisms between X and Y that send I to J, and others that send I to $-J$. It seems certain that $P(I)$ is either $P(J)$ or $P(-J)$ — there are no other options — but it also seems indeterminate which. A natural diagnosis is that it is actually indeterminate which place $P(I)$ is; it refers indeterminately to one of two possibilities. On the other hand, it *must* turn out that $P(I)$ and $P(-I)$ are distinct.[14] The reference of these terms is coordinated in a certain way.

In summary, let $S(X)$ and $S(Y)$ denote the structures of X and Y. $P(I)$ and $P(-I)$ are distinct elements, or places, of $S(X)$; $P(J)$ and $P(-J)$ are distinct elements, or places, of $S(Y)$. And although $S(X)$ equals $S(Y)$, there is no fact of the matter whether $P(I)$ equals $P(J)$.

The analogy to intensional typing should be clear. My suggestion is to take it as more than an analogy. I think it is the best hope for *explaining* what is going on in HoTT.

Let me explain the view in more detail. The idea is that reference to structures and their places goes by way of systems. Roughly speaking, a type symbol like $\ulcorner A \urcorner$ has as its *sense* or *intension* a system $\llcorner A \lrcorner$ (in the basic case, $\llcorner A \lrcorner$ is just a set). It has as its *referent* a structure A. More precisely, $\ulcorner A \urcorner$ refers to A under the guise 'the structure of $\llcorner A \lrcorner$'; in the earlier, more transparent notation, A is $S(\llcorner A \lrcorner)$. Similarly, a term symbol $\ulcorner a \urcorner$ 'of type A' has as its intension an element of $\llcorner A \lrcorner$, and as its reference a place in the structure A. More precisely, $\ulcorner a \urcorner$ refers to a under the guise 'the place

[14] This point in particular has been the subject of much debate; I will briefly consider that debate in ch. V.2.

occupied by $\llcorner a \lrcorner$'. The theory will be about structures like A and their places in the usual way that a discourse is about the referents of its words. In particular, equations will express sameness of referents. Still, the intensions of the words play their own role. On the present interpretation, one particular role they play is in *coordinating reference to places*.

To vary the example, consider the 'cardinality-two' structure which is exemplified by any two-element set (Shapiro, 2008). Let $\llcorner A \lrcorner$ and $\llcorner B \lrcorner$ be two such sets, with elements $\llcorner a \lrcorner, \llcorner a' \lrcorner$ and $\llcorner b \lrcorner, \llcorner b' \lrcorner$ respectively. So $\llcorner a \lrcorner, \llcorner a' \lrcorner$ occupy the two places of A, as do $\llcorner b \lrcorner, \llcorner b' \lrcorner$. Does $\llcorner a \lrcorner$ occupy the same place as $\llcorner b \lrcorner$ or the same place as $\llcorner b' \lrcorner$? In other words, does a equal b? As before, it seems to me that there is no determinate answer to this question. That, in short, is my explanation of why $\ulcorner a = b \urcorner$ is badly typed. The obvious further diagnosis is that there is no fact of the matter which place is occupied by $\llcorner a \lrcorner$. There are some alternative views in this vicinity. For example, following Breckenridge and Magidor (2012), perhaps there is a fact of the matter which place is a, but it is determined 'arbitrarily', making it effectively unknowable. On this view, it would be similarly arbitrary and epistemically indeterminable whether $a = b$. Here is why I prefer my view. We know in this situation what it takes to *settle* the truth values of identity claims: if we fix an isomorphism $\llcorner f \lrcorner$ between $\llcorner A \lrcorner$ and $\llcorner B \lrcorner$, then it is natural to say that, with respect to this isomorphism, $\llcorner a \lrcorner$ occupies the same place as $\llcorner f \lrcorner(\llcorner a \lrcorner)$, or that $\ulcorner a \urcorner$ and $\ulcorner f(a) \urcorner$ corefer. It wouldn't make sense to *first* settle the coreference facts arbitrarily, and then to *further* settle them in this way, with respect to an isomorphism. Rather, without such an isomorphism, the context just doesn't settle which terms corefer.

Admittedly, when I say that it is indeterminate whether $a = b$, and when I say there is no fact of the matter whether $a = b$, and so on, I am ambivalent about the best way to cash this out. However, I am thinking about it in the following way, which at least captures all the data. Suppose that $\llcorner A \lrcorner$ and $\llcorner B \lrcorner$ are isomorphic systems. On my way of thinking, *cohabitation* is a relation between $\llcorner a \lrcorner \in \llcorner A \lrcorner$, $\llcorner b \lrcorner \in \llcorner B \lrcorner$, and an isomorphism $\llcorner f \lrcorner : \llcorner A \lrcorner \to \llcorner B \lrcorner$. In other words, an equation $\ulcorner a = b \urcorner$ expressing the fact that $\llcorner a \lrcorner$ and $\llcorner b \lrcorner$ inhabit the same place only makes sense *relative to* an isomorphism $\llcorner f \lrcorner$. We should not expect a well-typed equation $\ulcorner a = b \urcorner$, but we should expect one roughly of the form $\ulcorner a =_f b \urcorner$. Of course, we do have a well-typed equation $\ulcorner f(a) = b \urcorner$, and, on my view, this captures the meaning of $\ulcorner a =_f b \urcorner$.

Finally, let me explain what I meant when I said that the intensions of terms coordinate reference to places. Although $\ulcorner a = b \urcorner$ is badly typed, $\ulcorner a = a' \urcorner$ is not. (I could also write the latter as $\ulcorner a =_{\mathrm{id}} a' \urcorner$.) It is determinate whether $\llcorner a \lrcorner$ and $\llcorner a' \lrcorner$ occupy distinct places in A, or, in other words, whether a and a' *are* distinct places. For that is just a matter of the distinctness of $\llcorner a \lrcorner$ and $\llcorner a' \lrcorner$. And this is what I mean when I say that the system $\llcorner A \lrcorner$ 'coordinates the reference' of $\ulcorner a \urcorner$ and $\ulcorner a' \urcorner$. I might stretch to say, more generally, that any isomorphism $\llcorner f \lrcorner : \llcorner A \lrcorner \to \llcorner B \lrcorner$ coordinates the reference of terms $\ulcorner a \urcorner$ and $\ulcorner b \urcorner$; then, as a special case, the identity function $\llcorner A \lrcorner \to \llcorner A \lrcorner$ coordinates the reference of $\ulcorner a \urcorner$ and $\ulcorner a' \urcorner$. At any rate, terms of syntactically the same type have their reference coordinated automatically, and terms of syntactically different types are coordinated with respect to isomorphisms. Remembering that isomorphisms are equivalent to elements of the constructive identity type $A \overset{**}{=} B$, this gives a new gloss on what constructive identity might be. An element of $A \overset{**}{=} B$ is a way of coordinating the intensions of $\ulcorner A \urcorner$ and $\ulcorner B \urcorner$.

Even if we can introduce enough context (the isomorphism $\llcorner f \lrcorner$) to settle whether $a = b$, this does not settle in any absolute sense the referents of $\ulcorner a \urcorner$ and $\ulcorner b \urcorner$ separately. It only coordinates them. Nothing I have said or can point to will determine which place of A is occupied by $\llcorner a \lrcorner$, assuming that $\llcorner A \lrcorner$ has automorphisms that affect $\llcorner a \lrcorner$. This is, I think, how it should be. Shapiro (2012) seems to agree that ordinary determinate reference to places is sometimes impossible. The thought is just that strong discernibility is necessary for the kind of reference in question. For example, consider the structure of a committee with two members plus a chair. Suppose that the two common members have the same powers and responsibilities, and are selected from a common pool. It would be pretty strange to ask which member occupied which position in the committee. There are two places on the committee, and they are occupied by the two people, but there is simply nothing more to say about which person occupies which place. In some situations, there may happen to be a natural bijection between the members of one committee and those of another. For example, maybe the Committee for Structuralism and the Committee for Category Theory have the same members. One might stretch to say that each member has the same place on the first committee as on the second. But, it seems to me, there is nothing about the structure of each committee, taken separately, that makes it so.

6.1. Other Structures. I have been talking in general terms about places in structures. But if $\ulcorner A \urcorner$ is a type-symbol, the structure A is just a cardinality structures, the structures of a set. I want to pause here to note how things work for more complicated structures. The discussion will importantly modify my view that the intension of a type-name is a particular set.

In section 2, I explained how all the structures of a certain kind can be gathered into a single 'structure-type', like the structure-type of complete ordered fields, or the structure-type of groups. I've discussed the univalence axiom that isomorphic types are equal. I also mentioned that the skeletal axiom for more complicated structures automatically follows: elements of a structure type that are isomorphic by the usual model-theoretic criteria come out equal.

To see why this is surprising, consider again the structure type of pointed sets. Again, we know that if $\llcorner A \lrcorner$ and $\llcorner B \lrcorner$ are isomorphic sets, and $\llcorner a \lrcorner$ and $\llcorner b \lrcorner$ are any elements of $\llcorner A \lrcorner$ and $\llcorner B \lrcorner$ respectively, then the pointed sets $(\llcorner A \lrcorner, \llcorner a \lrcorner)$ and $(\llcorner B \lrcorner, \llcorner b \lrcorner)$ are isomorphic. We would like to say that the structures of these pointed sets are therefore equal. How can we express this equality? Simply by the equation

$$(A, a) = (B, b). \tag{28}$$

At first, this looks strange. Usually, two ordered pairs are equal if and only if their corresponding components are equal. Certainly, as far as Zermelo theory is concerned, pointed sets $(\llcorner A \lrcorner, \llcorner a \lrcorner)$ and $(\llcorner B \lrcorner, \llcorner b \lrcorner)$ are equal if and only if $\llcorner A \lrcorner = \llcorner B \lrcorner$ and $\llcorner a \lrcorner = \llcorner b \lrcorner$. But we can't, and mustn't, analyse (28) as a conjunction of $A = B$ and $a = b$. That is exactly the kind of analysis that intensional typing rules out. What we *want* to say is that the structure of $(\llcorner A \lrcorner, \llcorner a \lrcorner)$ equals the structure of $(\llcorner B \lrcorner, \llcorner b \lrcorner)$. It seems most promising to say that the intension of $\ulcorner (A, a) \urcorner$ is the pointed set $(\llcorner A \lrcorner, \llcorner b \lrcorner)$; its referent is the structure of this pointed set. The formula (28) expresses equality between two such structures. A similar story goes for elements of other structure-types. (It may seem wrong to think of a structure-type as a collection of structures; shouldn't it be a structure itself? I will return to this point in Chapter V.1.)

As a special case, there is, as promised, a unique terminal object **1**; any two types have a unique product; and so on. However, this raises an interesting issue for my story about intensions. The referent of ⌜**1**⌝ is the structure of any singleton set; but what is the intension? According to my basic story, it should be some particular singleton set. Which one? Or, given types A and B, what is the intension of ⌜$A \times B$⌝? Presumably it is a product of $\llcorner A \lrcorner$ and $\llcorner B \lrcorner$; but which product is it? Perhaps the answer to each of these questions is 'Some arbitrary one.' But a more satisfying answer would be that **1** is the structure of any and all singleton sets, and $A \times B$ is the structure of any and all products of $\llcorner A \lrcorner$ and $\llcorner B \lrcorner$.

This answer isn't meant to be completely general. For example, if a type A has five elements, I still want to say that A is given to us as the structure of some particular five-element set, not as the structure of any and all five-element sets. The motivating idea, again, was that we can refer to places of A as the places occupied by elements of systems. To get determinate identities between terms of type A, we need the elements to belong to one and the same system.

In contrast, determinate identity between terms of type **1** or $A \times B$ does not require a particular singleton set or a particular product to be in the offing. If $\llcorner X \lrcorner$ and $\llcorner Y \lrcorner$ are singleton sets, with elements $\llcorner x \lrcorner$ and $\llcorner y \lrcorner$, it is completely determinate that $\llcorner x \lrcorner$ and $\llcorner y \lrcorner$ occupy the same place in the structure of any and all singleton sets. There is only one place they *could* occupy. There is no need to have one particular set coordinating the reference. Similarly, any two products of $\llcorner A \lrcorner$ and $\llcorner B \lrcorner$ are canonically isomorphic: there is a uniquely natural way to line up their elements. Thus, if $\llcorner x \lrcorner$ and $\llcorner y \lrcorner$ are elements of possibly distinct products of $\llcorner A \lrcorner$ and $\llcorner B \lrcorner$, it is completely determinate whether they occupy the same place in the structure $A \times B$ of each and every product of $\llcorner A \lrcorner$ and $\llcorner B \lrcorner$.

Finally, note that $A \times B$ is just a cardinality structure like any other. It comes with functions to A and to B. But these functions are not intrinsic to $A \times B$ itself. They depend on the fact that $A \times B$ is *given* to us in a certain way; they are determined by the intension of $\ulcorner A \times B \urcorner$. The more one looks at it, the more work the intensional data has to do. But that is not surprising. As made clear in ch. III.3, complicated expressions for types are analogous to sentences of predicate logic. Logic (and most of life) is only interesting because the true can be expressed in intensionally complicated ways.

7. Structuralism Without Equality?

I have articulated two views about equality in homotopy type theory. On the first view, the role of equality (at least, between sets) is played by non-propositional types $A \overset{**}{=} B$, and the resulting 'logic' of equality is intuitionistic. On the second view, the equality types $A = B$ are genuinely propositional. But only a weak version of the indiscernibility of identicals holds, in the sense that typing is intensional.

I have suggested an interpretation of intensional typing; but the reader might be tempted to conclude that *neither* $A \overset{**}{=} B$ nor $A = B$ really represents identity. In this section, I consider the merits of this view. The basic problem with it is that it leaves open the question of what the types $A \overset{**}{=} B$ and $A = B$ are actually about. I will argue that we cannot hope for some *third* account of identity and that, as far as HoTT goes, we should accept mere identity as the best account there is.

7.1. Judgmental Equality? I said just now that there can be no third account of identity in homotopy type theory. That is not quite right: we can at least use equality in a definitional sense, as I have done frequently, without comment, in this thesis. In homotopy type theory, this definitional role is officially played by so called *judgmental equality*, represented by the symbol $\equiv$. However, because judgmental equality is part of the formal system, one might hope that it could play a more than merely definitional role. I now argue that such hope is misplaced, although I would slightly prefer to say that judgmental equality expresses equality of intension. An equation $\ulcorner A \equiv B \urcorner$ does not directly express a relation between A and B, but only a relation between the intensions of $\ulcorner A \urcorner$ and $\ulcorner B \urcorner$. Still, if that is as far as judgmental equality goes, then we must look elsewhere for a full-blooded notion of equality.

Let us look more closely at how judgmental equality is typically used. If one wants to abbreviate an expression s by a new symbol α, then one can do so by stipulating a judgmental equality $\ulcorner \alpha \equiv s \urcorner$. That is an example of judgmental equality as definitional. However, the use of judgmental equality is somewhat more general than that. For example, consider how the rules (7), (8) of pair-formation and projection are related. We expect that, for any $a_1 : A_1$ and $a_2 : A_2$, $\mathrm{pr}_1(a_1, a_2)$ should equal a_1. In the standard treatment of function types, this equality is postulated as a judgmental equality

$$\mathrm{pr}_1(a_1, a_2) \equiv a_1.$$

Neither side of the equation is a straightforward abbreviation of the other. But it would be plausible to claim that the two sides have the same sense.

Here is the reason that judgmental equality cannot go beyond this sort of definitional or intensional use. Judgmental equality $A \equiv B$ is not a type. Remember that the logical functors were defined as operations on types. Thus one cannot apply logical functors to $A \equiv B$. Because of this, one cannot formally express the idea that A and B are *not* judgmentally equal, or that a type X has five judgmentally distinct elements, or anything like that. Even if we thought judgmental equality was full-blooded equality between referents, it would not be able to do the work that equality is ordinarily supposed to do in mathematics.

Instead of being a type, a judgmental equation $\ulcorner A \equiv B \urcorner$ has the same status as a typing declaration like $\ulcorner a : A \urcorner$. Each of these can figure as an input or an output in a construction. (They are both said to express 'judgments', but this traditional terminology seems unhelpful.) The basic rule of judgmental equality is that judgmentally equal terms can be substituted one for another. For example, here is a valid construction using judgmental equality:

$$\frac{a : A \quad A \equiv B}{a : B} \tag{29}$$

A term of type A is equally well of type B. Didn't this sort of reasoning lead to problems for the indiscernibility of isomorphs? The formal point is that, since we cannot apply logical functors to judgmental equalities, we cannot parlay (29) into a counterexample like (24). We cannot define the predicate $F(X, x)$ by '$X \equiv \mathbb{N}$ *and*...'. This observation corroborates my view that judgmental equality expresses equality of intension. 'The morning star' and 'the evening star' have different intensions, but we cannot parlay this into a genuine counterexample to the indiscernibility of identicals.

7.2. What If Not Equality? We must leave aside judgmental equality. The view on offer is that neither constructive equality nor its propositional truncation nor judgmental equality expresses equality of referent.

It is clear this position must be qualified. We cannot discount typical equality in all cases. We do need identity predicates to do mathematics. We need to be able to express the fact that two plus two is *not* equal to five, and judgmental equality won't suffice for this: it can't be negated. However, a basic observation from categorical set theory was that we *don't* ordinarily need equations between sets. That is why the omission of object-identities in categorical set theory was pragmatically acceptable. The qualified position, then, is that we accept as genuine some form of typical equality between elements of sets, but not between sets themselves. Spelling out this qualified position is bound to be a little delicate. Within the larger universe $\mathcal{U}'$ there will be some objects (the sets) between whose elements typical equality counts as genuine equality, and other objects (like $\mathcal{U}$) for which some other interpretation is needed.

Still, suppose the line can be drawn between one kind of object and the other. Such a move would seem to undo the advantages that HoTT has over categorical set theory as described in Chapter II. In giving up equations between sets, we would have to give up the skeletal axiom (Univalence would not be an axiom about *identity*), and we would fall back into an account of 'structural properties' rather than 'structural objects'. In particular, we would not have a conventional 'unique reference' account of the use of singular terms; we would have to fall back on an account like those I described in ch. III.1.2. Finally, the position leaves open the most interesting questions about the interpretation of HoTT. How *should* we interpret typical equality in cases where it isn't genuine equality? How should we interpret Univalence?

Two slightly different answers suggest themselves. One answer is that an 'equation' between types expresses not *equality* but *indiscernibility*. The other answer is that an 'equation' expresses not equality per se but *equality of structure*.

Let me look more closely at this second answer before I explain the basic problem that faces both of them. On this reading, $\ulcorner A \urcorner$ and $\ulcorner B \urcorner$ refer to systems A and B, but $=$ expresses sameness of structure rather than identity of systems. Of course, we are used to the idea that *isomorphism* means sameness of structure. So, on this reading, Univalence is something of a tautology. However, it does have an effect: it transfers to isomorphism the properties postulated of $=$. The main such property is the identity of indiscernibles (or its generalisation in the path induction rule). So the effect of Univalence is to postulate the indiscernibility of of isomorphs:[15] for $A, B : \mathcal{U}$ and $P : \mathcal{U} \to \mathcal{U}'$,

(30)
$$\frac{f : \mathrm{Iso}(A, B)}{f_* : P(A) \to P(B)}$$

Here I have written it in its general constructive form. Even if we eliminate cosntructive and mere equality, we must still attend to the distinction between (as it were) constructive and mere isomorphism.

Now, we already had in Chapter II a theorem about the indiscernibility of isomorphs in categorical set theory. What is new here? The earlier theorem was really a metatheorem about what kinds of properties one could define in the first-order theory of categories. The work being done

[15]A similar point is made more formally by Pelayo and Warren (2012, pp. 40–41).

by Univalence is to express indiscernibility of the isomorphs *within* the formal system. To see what this means, consider what would happen if we tried to *add* a new, 'genuine' identity predicate to the theory. We would want, for each pair of sets A and B, a proposition $I(A,B)$. Presumably, we want $I(A,A)$ to be true for each type A. But (30) would then yield that $A \cong B$ implies $I(A,B)$.[16] In short, genuine identity is entailed by isomorphism. So even if we deny that HoTT *as it stands* has identity predicates for types, any attempt to introduce such predicates leads inevitably back to the motto that 'isomorphic objects are equal'. In contrast, categorical set theory is compatible with the introduction of a genuine identity predicate that is not entailed by isomorphism — the usual identity predicate of Zermelo set theory.

But wait, there's more. The argument so far shows that $A \cong B$ implies $I(A,B)$. By Univalence, we could equally well say that $A = B$ implies $I(A,B)$. If genuine identity I satisfies at least the weak version of the indiscernibility of identicals, then we will also find the converse: $I(A,B)$ implies $A = B$. So genuine identity is logically equivalent to *whatever* is expressed by $=$. In fact, it then follows[17] that

genuine identity is genuinely identical to $=$.

This argument works whether we try to interpret $=$ as indiscernibility or as sameness of structure, or as something else. It is a version of the

[16]Explicitly: define $P(X) :\equiv I(A,X)$. Then (30) yields $I(A,A) \to I(A,B)$. Since $I(A,A)$ is true, we can conclude that $I(A,B)$ is true.

[17]In detail: for arbitrary $A,B : \mathcal{U}$, we have $I(A,B) \longleftrightarrow A \cong B$. Since both $I(A,B)$ and $A \cong B$ are propositions, the two functions involved in this biconditional must be inverse isomorphisms. Since isomorphism implies genuine identity, $I(A,B)$ is genuinely identical to $A \cong B$. (The reader may then use Function Extensionality (21) to deduce that I and $=$ are genuinely identical as functions.)

familiar argument that, given the resources of second-order logic, we can define identity in terms of indiscernibility.

In conclusion, it is logically possible to deny that $=$ expresses genuine equality. But if we do so, we must accept that there can be no *other* genuine equality predicate in HoTT. Such a position is not without precedent. In Resnik's view (Resnik, 1997, p. 210), 'patterns' are subject only to 'congruence' and not identity. In an older precedent, Frege denied that identity applies to concepts.[18] Both of these views deny that the entities in question are truly objects. It would be interesting to investigate these views more closely in the current context. But, on the face of it, the conceptual price of giving up equality seems greater to me than the price of simply accepting that it is expressed by mere identity. The price of the latter is intensional typing; I have tried to explain how it might be paid.

8. Conclusion

The Univalence Axiom amounts to a quite general skeletal axiom; it provides hope for homotopy type theory as an object-structuralist foundation. But the theory is also haunted by the spectre of intuitionistic logic. I explored how one might give a sympathetic interpretation of HoTT while preserving the idea that identity is a genuine relation, satisfying classical logic. For this narrow purpose, the main issue is intensional typing; I suggested how it might be explained in terms of a referential indeterminacy that naturally arises from object structuralism. In this context, the notion of constructive identity can be eliminated in favour of isomorphism. But I left wide open its general interpretation.

---

[18]Cf. Furth's introduction to (Frege, 1893), p. xliv.

CHAPTER V

Challenges

In this chapter I raise some issues that will have to be addressed in a fuller development of the ideas in this thesis.

1. Inter-Structure Identity

So far, I have generally taken the view that a structure is specified by an object in a category. Perhaps this doesn't always work; it is probably not true that an arbitrary object in an arbitrary category has a structure in any useful sense. But if we look at the category of the models of some theory — the category of sets, or pointed sets, or groups — then it is pretty natural to say that each object has a certain structure, and that isomorphic objects have the same structure, and that non-isomorphic objects do not. Univalence builds on this picture. We have a structure-type PSet of pointed sets (for example), in which the objects are individuated exactly by isomorphism. Similarly, we have a structure-type of groups, and of course a structure-type $\mathscr{U}$ of sets (see ch. III.1.) Univalence works well in all these cases.

However, there are a number of puzzles about this picture which ought to be addressed if we are to have a full account of HoTT as a theory of structure. I will introduce some of them now, although a full investigation will have to wait for another occasion. The basic question to be answered is: when does object A of structure-type $\mathscr{C}$ have the same structure as object B of structure-type $\mathscr{D}$? At first glance, Univalence goes a long way towards providing an answer. I will explain why things are not so simple, and then speculate about what the general picture might be.

First, why might Univalence help? To give a simple but not quite trivial example, consider the structure type

$$\mathrm{PSet} :\equiv \coprod_{(X:\mathscr{U})} \coprod_{(x:X)} \top$$

of pointed sets, and the structure type

$$\mathrm{PSet}' :\equiv \coprod_{(X:\mathscr{U})} \coprod_{(x:X)} \coprod_{(y:X)} x = y$$

of two-pointed sets whose two points are equal; call the latter shmointed sets. It is very tempting to think that pointed set structures and shmointed set structures are the same thing. In particular, it is very tempting to think that the pointed set (X, x) has the same structure as the shmointed set (X, x, x). Univalence appears to respect this intuition. At least, the function $(X, x) \mapsto (X, x, x)$ from pointed set structures to schmointed set structures is an isomorphism. If Univalence applies, then PSet equals PSet$'$. Pointed set structures are shmointed set structures!

However, unfettered, this kind of reasoning is catastrophic. Here is a dramatic example. Consider the structure type $\mathscr{N}$ of models of the second-order Peano axioms. Consider also the structure type $\mathscr{R}$ of complete ordered fields. Now, up to unique isomorphism, there is only one model of the second-order Peano axioms, and only one complete ordered field (the real numbers). This means that $\mathscr{N}$ and $\mathscr{R}$ are both one-element sets. If Univalence applies, $\mathscr{N}$ equals $\mathscr{R}$. Natural numbers structures are real numbers structures! This is, of course, absurd.

It might appear that we have to block this second example at all costs. Let me explain why one obvious way to do this is unattactive. We could deny Univalence for $\mathscr{U}'$. I am taking it for granted that the universe $\mathscr{U}$ satisfies Univalence, but we are not necessarily committed to Univalence for the larger universe $\mathscr{U}'$. If we denied it, then we could not infer $\mathscr{N} = \mathscr{R}$. More generally, we would not be committed to the view that elements of $\mathscr{U}'$ are cardinality structures. However, this move is unsatisfactory. First, it leaves open what the correct account of identity in $\mathscr{U}'$ is. Second, the original motivation for having a hierarchy of universes was that $\mathscr{U}'$ contains sets (or structures) that are 'too big' to fit in $\mathscr{U}$. But if the difference between $\mathscr{U}$ and $\mathscr{U}'$ is simply one of size, then the motivation for assuming Univalence for $\mathscr{U}$ applies just as well in the case of $\mathscr{U}'$.

In fact, I do not think that we have to block the example. Here is a diagnosis of what is going wrong. In some contexts, '$\mathscr{N}$' denotes a collection extensionally defined: the collection of natural-numbers structures. In other contexts — when we speak of $\mathscr{N}$ as an element of $\mathscr{U}'$ — it denotes a structure: the *structure of* the collection of natural numbers structures. The latter is the context in which $\mathscr{N} = \mathscr{R}$, this being an equation between elements of $\mathscr{U}'$. The equation says (a bit roughly) that the number of natural-numbers structures equals the number of real-numbers structures. That is as it should be, but we obviously cannot deduce that natural-numbers structures are real-numbers structures.

Strictly speaking, the same disambiguation must apply to the symbol '$\mathscr{U}$'. Sometimes '$\mathscr{U}$' denotes an extensionally given collection: the collection — rename it $\mathscr{S}$ — of all small cardinality structures. In this context, a term A of type $\mathscr{U}$ denotes one of those structures, and a term of type A denotes a place in that structure. But when we speak of $\mathscr{U}$ as an element of a larger universe $\mathscr{U}'$, '$\mathscr{U}$' denotes the *structure* of $\mathscr{S}$. In this second picture, a term A of type $\mathscr{U}$ denotes the place in that structure occupied by a certain small cardinality structure. This equivocation about the denotation of terms of type $\mathscr{U}$ does not lead to formal contradiction precisely because small cardinality structures are individuated in the same way as the places they occupy.

This kind of disambiguation prevents the catastrophic conclusion that real-numbers structurs are natural-numbers structures. But it also shows that we cannot deduce anything useful about the relationship between pointed set structures and shmointed set structures just from the equation $\mathrm{PSet} = \mathrm{PSet}'$. Thus the question with which I started remains untouched: when are two structure-types the same (as structure-types, rather than as cardinality structures) and when are elements of different structure-types the same as structures?

If we really are thinking of structure-types extensionally as collections of structures, then I suspect that homotopy type theory has little to say on its own about this question. The reason for scepticism is that we are asking whether two structure-types are *extensionally* the same – whether they contain the same structures — whereas Univalence is about whether they are (in one sense or another) *structurally* the same.

That isn't meant to be a knock-down objection. In the first case, I pointed out in ch. II.4.1 that the extension of a Zermelo set can be conceptualised structurally. More generally, there is an idea extant in the structuralist literature that (for example) to be a collection of structures of a certain kind isn't an extensional property at all; it just is to have a structure of a certain kind. I will call this idea *meta-structuralism*. To give a simple example, each structure-type $\mathscr{T}$ comes with a 'first projection' function $\mathrm{pr}_1 : \mathscr{T} \to \mathscr{U}$, mapping the structure of a system to the structure of the underlying set. Although $\mathscr{N}$ and $\mathscr{R}$ are isomorphic, there is no isomorphism between them that commutes with the first projections. To put it another way, HoTT distinguishes the structure of $(\mathscr{N}, \mathrm{pr}_1)$ from the structure of $(\mathscr{R}, \mathrm{pr}_1)$. This may be just what we need to distinguish natural numbers structures from real numbers structures.

This kind of meta-structuralist view may be latent in Awodey's 'top-down' structuralism (Awodey, 2004), but its clearest formal development may be the theory of the Category of Categories as a foundation (CCAF). This theory was seen even by Lawvere (1964) as more promising a foundation than categorical set theory. For some philosophical discussion of it, see (Hellman, 2006; Linnebo and Pettigrew, 2011). The meta-structuralist idea deserves more attention than I can give it here. In particular, it would be interesting to understand whether there is a formalism that stands in the same relation to CCAF that HoTT stands in to categorical set theory.

Let me conclude this tentative discussion by considering an example that illustrates the present concerns and raises a number of problems for further investigation. Let $o\mathscr{S}et$ be the structure-type of ordered sets. Consider its elements $(\mathbb{N}, \geq)$ and $(\mathbb{N}, \leq)$. Do they have the same structure? On the one hand, it is hard to escape the feeling that the difference between $\geq$ and $\leq$ is purely orthographic — a matter of the order in which one *writes* the arguments — and in no way a matter of metaphysics. For some very natural sense of the word 'structure', these systems undoubtedly have the same structure. On the other hand, as long as we are talking about mathematical structuralism, we should attend to the ways in which *mathematics* individuates structures. The fact remains that $(\mathbb{N}, \geq)$ and $(\mathbb{N}, \leq)$ are non-isomorphic as ordered sets,[1] and that must point to a legitimate sense in which their structures are distinct. It is noteworthy that $(\mathbb{N}, \geq)$ and $(\mathbb{N}, \leq)$ are 'definitionally equivalent' by the usual criteria (Hodges, 1997, p.54). So the example tells against the very common thought that definitionally equivalent systems, in general, have the same structure (e.g. Shapiro (1997, p. 91), following Resnik (1981), and Halvorson (2012)).

What light might meta-structuralism shed on this example? According to meta-structuralism, $o\mathscr{S}et$ (along with some relevant data) is an ante rem structure and $(\mathbb{N}, \geq)$ and $(\mathbb{N}, \leq)$ are places in this structure. This is analogous to the way in which 'the complex numbers' is supposed to be an ante-rem structure, and i and $-i$ are places in that structure. There is the well-known puzzle about what distinguishes i from $-i$ (of which more in section 2). Although it is certain that i and $-i$ are distinct, there is no further fact about *which* square root of -1 is i and which is $-i$; one can only say that i is i and $-i$ is $-i$. In IV.6 I put this down to indeterminacy of reference. If I was right, the relationship between $(\mathbb{N}, \geq)$ and $(\mathbb{N}, \leq)$ may be of fundamentally the same sort. Although '$(\mathbb{N}, \geq)$' and '$(\mathbb{N}, \leq)$' refer to determinately distinct structures, it is indeterminate which one refers to which.

Finally, let me point out a basic problem that this example raises for the structuralist credentials of homotopy type theory. I insisted that $(\mathbb{N}, \geq)$ and $(\mathbb{N}, \leq)$ have different structures, because they are not isomorphic as ordered sets. This stance does not require us to abandon the idea that $(\mathbb{N}, \geq)$ and $(\mathbb{N}, \leq)$ have the same structure in *some* sense of the word. And we can even explicate that sense by introducing a category in which these systems

[1]They are non-isomorphic, because, for example, $\exists(x : \mathbb{N}).\forall(y : \mathbb{N}).(yRx)$ is true when R is $\geq$ and false when R is $\leq$.

are isomorphic. For example, we could re-define a morphism between ordered sets (X,R) and (X',R') to be a function $f : X \to X'$ that is *either* order-preserving or order-reversing:

$$\begin{aligned} &\forall(x,y:X).(xRy \to f(x)R'f(y)) \\ &\vee\forall(x,y:X).(xRy \to f(y)R'f(x)). \end{aligned} \tag{31}$$

But this only shows that the notion of isomorphism is flexible enough to accommodate many different notions of structure. It does not show that the notion of structure implicit in (31) is the 'right' notion, and, if anything, mathematical practice points the other way. Here is the problem for the structuralist interpretation of homotopy type theory: although $(\mathbb{N},\geq)$ and $(\mathbb{N},\leq)$ are isomorphic in the sense of (31), one can't use Univalence to deduce that they are equal. That's because Univalence coordinates equality with the *standard* notion of isomorphism between ordered sets.[2] This is a serious qualification to Awodey's claim that 'two mathematical objects are identical if and only if they have the same structure' (Awodey, 2014). I discuss an even more important example in the next section.

2. Homotopy and Identity

A crucial feature of homotopy type theory that I have largely bracketed until now is the way it generalises set-theoretic mathematics. I have explained how it does this formally: there are cases in which propositional equality does not coincide with mere equality, and this has an impact (for example) on how functions are individuated (ch. IV.4). A fuller evaluation

[2]In the terminology of the 'Structure Identity Principle' of (*UFP*, ch. 9), (31) does not define a 'standard notion of structure'. The method of Rezk completion (*UFP*, ch 9) may allow us to overcome this problem, but, anyway, the point is that Univalence does not do it on its own.

of HoTT than I can provide here would attend more carefully to this generalisation. In this section, I make some preliminary comments about how this generalisation might intersect with the structuralist interpretation.

It may be useful to begin by recalling an old debate. One of the main preoccupations of the literature on ante rem structuralism has been the sense, if any, in which the places of an ante rem structure are individuated by or grounded in or otherwise metaphysically dependent upon the structure. Here is a stripped-down version of the issue. On the one hand, the complex numbers i and $-i$ are distinct; but, on the other, they seem to play the same role in the structure of the complex numbers, since they are related by an automorphism. Therefore the way complex numbers are individuated is different from the way in which places in the structure are individuated. Therefore the complex numbers can't *be* the places in that structure. This contradicts ante rem structuralism.

The obvious response is that the places i and $-i$ are 'the same' in being indiscernible, not in being equal. But then one might wonder what it is that makes them distinct. Ladyman (2005) tried to explain it in terms of 'weak discernibility', but further discussion (Ketland, 2006; Leitgeb and Ladyman, 2008; Shapiro, 2008) has shown this is not enough. Something like a consensus has developed that, indeed, ante rem structuralists must accept the distinctness of the places as a primitive feature of the structure.[3] This is, too, is the bottom line in the structuralist interpretation of homotopy type theory. The theory has nothing to say about what makes elements of the same object distinct.

[3]Whether or not this is a *problem* is more controversial: see MacBride (2005) and Shapiro's exchange with Keränen in (MacBride, 2006).

However, I think one can say a *little* more than this, from the point of view that 'structure is what is preserved by isomorphism'. Any isomorphism between sets — or between groups, or pointed sets, or complete ordered fields — is a bijection; it particular, it preserves the distinctness of the elements. That distinctness is, therefore, part of the structure. Identity is certainly special in that it is not schematic in the way other parts of the structure are. In different exemplifications of the natural numbers structure, the 'greater than' role is filled by different relations: in one model, it might be 'greener than', and, in another, 'closer to Betelgeuse than'. Identity is always identity; but it is part of the structure nonetheless.

Non-Structural Distinctness. This observation that isomorphisms preserve the distinctness of elements may seem vacuous. But, far from being vacuous, it is not always true. I will sketch two of the central examples.

The first example is that of *homotopy structure*. In the basic version of the story, the things that exemplify homotopy structure are topological spaces. Roughly speaking, two spaces exemplify the same homotopy structure if they can be deformed into each other by arbitrary continuous stretching and shrinking. As basic examples, a solid ball has the same homotopy structure as a disk, a line, or a single point. A two-handled mug has a different homotopy structure from those, but the same homotopy structure as the figure '8'. A hollow sphere has yet another homotopy structure. The notion of 'isomorphism' that corresponds to this notion of structure is called *homotopy equivalence*. It differs from the standard notion of 'isomorphism between topological spaces', namely, homeomorphism. The exact definition isn't important for the limited discussion I have space for here. What should already be evident is that homotopy equivalences need not be bijections; they need not preserve the distinctness of points. For example, a solid ball is homotopy-equivalent to a single point, but the ball and the point have different cardinalities. In this particular case, *any* function between them will determine a homotopy equivalence.[4]

A second example is given by category theory. For simplicity, think about the standard two-sorted theory of categories (ch. II.1). This theory is stated in a standard first-order language, and there is a standard notion of isomorphism between models of such a theory. However, it is generally recognised that the structure that category theory is typically intended to capture is reflected in a more general kind of 'isomorphism', namely the so-called equivalences of categories. These equivalences need not preserve the distinctness of objects. Again, equivalent categories may even have different cardinalities of objects. This observation provides another way of getting at the issues raised in Chapter II. Equations between objects in a category are suspect precisely because they are not invariant under equivalence. They do not reflect anything intrinsic to the structure of the category.

These examples raise serious questions about the cogency of ante rem structuralism. Do category structures have places? Do homotopy structures? If the homotopy structure of a ball has places, how many does it have? Uncountably many, like the ball, or just one, like the point? To put it optimistically: if these structures have places, then there is something novel about the way in which those places are individuated.

The Homotopy Interpretation. I have several times mentioned that one motivation for homotopy type theory is the desire for a synthetic

[4]The reader may wonder how it is that an isomorphism can fail to be a bijection, since an isomorphism must be invertible, and, among functions, only bijections are invertible. A proper answer to this question would lead us deep into the heart of the matter. For now, the short answer is that a morphism in the relevant category is not a function at all, but an equivalence class of functions. It is still true that any function between the ball and the point *determines* a homotopy equivalence.

treatment of homotopy theory, distinct from its usual development in set theory. The exact motivation is bound to be complicated, but we can now grasp something about it. Each set-theoretic exemplification of a homotopy structure has properties that are not part of the structure. For example, each model has some particular cardinality. Of course, if we are talking about models in Zermelo set theory, then Benacerraf's observation automatically applies: the extension of the model is not part of the structure. But the *new* problem is a problem even if we look for models among the sets of categorical set theory or indeed those of homotopy type theory. In these theories, we cannot distinguish equinumerous sets by their extensions, but we can still distinguish sets that aren't equinumerous. Univalence does *not* entail that isomorphic models of homotopy structure are equal, if these models are constructed by the usual set-theoretic methods.[5] It is for roughly this reason that set theory *can't* yield a fully perspicacious explication of homotopy structure. Homotopy type theory, in generalising set theory, seeks a better way forward.[6]

The position, then, is this. Examples like homotopy theory and category theory raise serious problems for ante rem structuralism. But the same kind of problems lie behind HoTT's attempt to generalise set theory. If the set-theoretic part of HoTT does, indeed, articulate a structuralist vision of mathematics, then perhaps HoTT as a whole will be faithful guide to what structuralists can and should say about these more difficult examples. There is, of course, the obverse risk: perhaps an investigation of the

[5]At the end of the previous section I gave a simpler example in which Univalence fails to entail that 'isomorphic systems are equal'. That example is worth further analysis, but it is not as central to contemporary mathematics as homotopy theory and category theory are.

[6]I made a few further remarks about the homotopy interpretation in IV.4.

larger picture will vitiate the kinds of structuralist interpretations that I have suggested here.

CHAPTER VI

Summing Up

I began this thesis with a very conservative view of structuralist foundations. 'To get the ball rolling,' I wrote, 'I am happy to grant that Zermelo set theory is true and epistemically secure.' Let us see where that ball has rolled.

On the conservative view, the *two-sorted* version of categorical set theory is a property-structuralist theory of *sets*. Its objects are Zermelo sets, but it focuses only on their structural properties — at least as far as monadic predicates go. As soon as we consider more complicated predicates, or, what amounts to the same thing, monadic predicates of more complicated systems, it ceases to be a property-structuralist theory. The language can express non-structural properties of spans, of pointed sets, and on and on. There is room for a more progressive view, according to which the sets of CST are not Zermelo sets, but something sui generis, either abstracted from Zermelo sets or corresponding to some entirely new conception. These sui generis sets would have only structural properties; we would have an object-structuralist set theory. But the same problem would arise: we would not have a structuralist foundation for mathematics in general.

Either way, the problem can be 'fixed' by eschewing set-identity predicates. This required adopting strict type distinctions into the language. It is hard to understand this as more than a linguistic gambit, a means for *ignoring* non-structural properties. So while categorical set theory has some merit as a *property*-structuralist foundation for mathematics in general, it does not seem promising as an articulation of *object* structuralism. *As* a property-structuralist foundation, its objects might be systems constructed either from Zermelo sets (on the conservative reading) or from sui generis categorical sets (on the progressive one).

Homotopy type theory, in contrast, while building conceptually on categorical set theory, looks like a candidate for an object-structuralist foundation. The Univalence axiom ensures that isomorphic systems are identical. It is worth pointing out, again, that this implication of Univalence is not completely general. But, as far as it goes, I suggested an object-structuralist gloss on Univalence: ante rem structures with isomorphic *systems of places* are identical. However, to connect back to all the problems raised in Chapter II, this theory of ante rem structures is not completely autonomous. At least, we need an explanation of intensional typing; the one I gave appealed to the idea that structures are given to us as the structures of systems. These systems might, again, be constructed from Zermelo sets or from sui generis categorical sets. Given the way in which homotopy type theory builds on categorical set theory, the latter option seems particularly attractive; but I am afraid I cannot say much about it either way. We could even look further afield; the systems could include systems of concreta, or, for all I've said, 'free creations of the human mind.' *Some* systems are needed, though — or we need another account of intensional typing.

In short, there are a range of possibilities within the structuralist genre. More work will be required to see how well they hang together. What I have had time and space for in this thesis can only be the beginning of the conversation. Much more attention should be paid, in particular, to the ways in which HOTT seeks to generalise set-theoretical mathematics. This generalisation should be philosophically interesting in its own right, if it provides a perspicacious treatment of homotopy theory and its kin. But it is also bound to modify, enlighten, or destroy the kind of possibilities that I have offered here.

Bibliography

Awodey, S. (2004). An answer to Hellman's question: 'Does category theory provide a framework for mathematical structuralism?'. *Philosophia Mathematica*, 12(1).

Awodey, S. (2009). From sets to types to categories to sets. *Carnegie Mellon University Research Showcase*.

Awodey, S. (2014). Structuralism, invariance, and univalence. *Philosophia Mathematica*, pages 1–11.

Benacerraf, P. (1965). What numbers could not be. *The Philosophical Review*, pages 47–73.

Boolos, G. (1971). The iterative conception of set. *Journal of Philosophy*, 68(8):215–231.

Breckenridge, W. and Magidor, O. (2012). Arbitrary reference. *Philosophical studies*, 158(3):377–400.

Burgess, J. P. (1999). Book review: Stewart Shapiro. philosophy of mathematics: Structure and ontology. *Notre Dame Journal of Formal Logic*, 40(2):283–291.

Cole, J. C. (2010). Mathematical structuralism today. *Philosophy Compass*, 5(8):689–699.

Evans, G. (1978). Can there be vague objects? *Analysis*, page 208.

Feferman, S. (2004). Typical ambiguity: trying to have your cake and eat it too. In Link, G., editor, *One Hundred Years of Russell's Paradox*, pages 135–152.

Frege, G. (1964/1893). *The Basic Laws of Arithmetic*. University of California Press.

Halvorson, H. (2012). What scientific theories could not be. *Philosophy of science*, 79(2):183–206.

Hellman, G. (2003). Does category theory provide a framework for mathematical structuralism? *Philosophia Mathematica*, 11(2):129–157.

Hellman, G. (2006). What is categorical structuralism? In *The Age of Alternative Logics*, pages 151–161. Springer.

Hodges, W. (1997). *A Shorter Model Theory*. Cambridge University Press.

Kapulkin, C., Lumsdaine, P. L., and Voevodsky, V. (2012). The simplicial model of univalent foundations. *arXiv preprint arXiv:1211.2851*.

Keränen, J. (2006). The identity problem for realist structuralism II : A reply to shapiro. In MacBride, F., editor, *Identity and Modality*, pages 146–163. Oxford University Press.

Ketland, J. (2006). Structuralism and the identity of indiscernibles. *Analysis*, 66(292):303–315.

Ladyman, J. (2005). Mathematical structuralism and the identity of indiscernibles. *Analysis*, 65(287):218–221.

Lawvere, F. W. (1994). Cohesive toposes and cantor's' lauter einsen'. *Philosophia Mathematica*, 2(1):5–15.

Lawvere, F. W. (2005/1964). An elementary theory of the category of sets (long version) with commentary. *Repr. Theory Appl. Categ*, 11:1–35.

Leitgeb, H. and Ladyman, J. (2008). Criteria of identity and structuralist ontology. *Philosophia Mathematica*, 16(3):388–396.

Lewis, D. (1988). Vague identity: Evans misunderstood. *Analysis*, pages 128–130.

Linnebo, Ø. and Pettigrew, R. (2011). Category theory as an autonomous foundation. *Philosophia Mathematica*, 19(3):227–254.

Mac Lane, S. and Moerdijk, I. (1997). *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*. Cambridge University Press.

MacBride, F. (2005). Structuralism reconsidered. In *The Oxford Handbook of Philosophy of Mathematics and Logic*, pages 563–589. Oxford University Press.

MacBride, F. (2006). *Identity and Modality*. Oxford University Press.

Makkai, M. (1995). First order logic with dependent sorts, with applications to category theory. *Unpublished*.

Martin-Löf, P. (1984). *Intuitionistic Type Theory: Notes by Giovanni Sambin of a series of lectures given in Padova, June 1980*.

McLarty, C. (1993). Numbers can be just what they have to. *Noûs*, 27(4):487–498.

McLarty, C. (2004). Exploring categorical structuralism. *Philosophia Mathematica*, 12(1):37–53.

McLarty, C. (2005). Learning from questions on categorical foundations. *Philosophia Mathematica*, 13(1):44–60.

Parsons, C. (1990). The structuralist view of mathematical objects. *Synthese*, 84(3):303–346.

Pelayo, Á. and Warren, M. A. (2012). Homotopy type theory and Voevodsky's univalent foundations. *arXiv preprint arXiv:1210.5658*.

Reck, E. H. and Price, M. P. (2000). Structures and structuralism in contemporary philosophy of mathematics. *Synthese*, 125(3):341–383.

Resnik, M. D. (1981). Mathematics as a science of patterns: Ontology and reference. *Nous*, pages 529–550.

Resnik, M. D. (1997). *Mathematics as a Science of Patterns*. Oxford University Press.

Shapiro, S. (1997). *Philosophy of mathematics: Structure and ontology*. Oxford University Press.

Shapiro, S. (2005). Categories, structures, and the Frege-Hilbert controversy: The status of meta-mathematics. *Philosophia Mathematica*, 13(1):61–77.

Shapiro, S. (2006). Structure and identity. In MacBride, F., editor, *Identity and Modality*, pages 34–69. Oxford University Press.

Shapiro, S. (2008). Identity, indiscernibility, and ante rem structuralism: The tale of i and $-i$. *Philosophia Mathematica*, 16(3):285–309.

Shapiro, S. (2012). An "i" for an i: Singular terms, uniqueness, and reference. *The Review of Symbolic Logic*, 5(3):380–415.

Univalent Foundations Program (2013). *Homotopy Type Theory: Univalent Foundations of Mathematics*. http://homotopytypetheory.org/book, Institute for Advanced Study.